

UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION

THE STATE OF ALABAMA, *et al.*

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF
COMMERCE, *et al.*,

Defendants.

No. 3:21-cv-00211-RAH-ECM-KCN

DEFENDANTS' RESPONSE IN OPPOSITION TO
PLAINTIFFS' MOTION FOR PRELIMINARY INJUNCTION
AND PETITION FOR WRIT OF MANDAMUS

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND.....	4
A. The Decennial Census	4
B. The Census Act’s Confidentiality Provisions	5
C. The Rise of Computing Power and Its Implications for Confidentiality	6
D. Differential Privacy	8
E. The Census Bureau’s Delivery of Redistricting Data	16
ARGUMENT.....	19
I. Plaintiffs Lack Standing.....	19
A. Plaintiffs Have Not Sustained Any Injuries-in-Fact	20
1. Plaintiffs Are Not Injured by Differential Privacy	20
a. Informational Injury	20
b. Sovereign Injury	29
c. Federal Funding	33
d. Vote Dilution	35
e. Section 209.....	36
2. Plaintiffs Are Not Injured by Delayed Redistricting Data	36
B. Plaintiffs’ Alleged Injuries Are Not Traceable to Defendants’ Actions	38
1. Plaintiffs’ Alleged Injuries Cannot Be Traced to Defendants’ Plan to Use Differential Privacy	38
2. Plaintiffs’ Alleged Injuries Cannot Be Traced to Defendants’ Delay in Producing Redistricting Data	40

C.	Plaintiffs’ Purported Injuries Are Not Redressable	40
1.	Enjoining Differential Privacy Would Not Redress Plaintiffs’ Alleged Injuries	41
2.	Requiring the Census Bureau to Produce Redistricting Data Sooner Would Not Redress Plaintiffs’ Alleged Injuries	42
II.	Plaintiffs Are Not Entitled to a Preliminary Injunction.	44
A.	Plaintiffs Are Unlikely to Succeed on the Merits of Their Differential Privacy Claims.....	45
1.	Plaintiffs’ Census Act Claim Is Not Likely to Succeed	45
2.	The Individual Plaintiffs’ Equal Protection Claim Is Not Likely to Succeed	48
3.	Plaintiffs’ APA Challenges to Differential Privacy Are Not Likely To Succeed.....	49
a.	The Differential Privacy Announcement Was Not Final Agency Action	49
b.	Even Assuming the Differential Privacy Announcement Constituted Final Agency Action, It Did Not Violate the APA	55
4.	The Doctrine of Laches Bars Plaintiffs’ Differential Privacy Claims	59
B.	Plaintiffs’ Challenge to the February 12 Press Release Is Not Likely to Succeed.	61
1.	Plaintiffs’ Claim that the Press Release “Violates the Census Act” Is Not Likely to Succeed.....	61
2.	Alabama’s APA Challenge to the February 12 Press Release Is Not Likely to Succeed.....	64
a.	The February 12 Press Release Was Not Final Agency Action.....	64
b.	The February 12 Press Release is Not Arbitrary or Capricious	66

C.	Plaintiffs Will Suffer No Harm, Much Less Irreparable Harm.	68
1.	Plaintiffs Have Not Established Irreparable Harm Due to Differential Privacy	68
2.	Plaintiffs Have Not Established Irreparable Harm on Their Delay Claim.....	71
D.	Defendants and the Public Would Be Harmed by an Injunction.	73
III.	Mandamus Relief Is Unavailable.	75
CONCLUSION.....		78

INTRODUCTION

Every decade, the United States Census Bureau has the responsibility of “counting the whole number of persons in each State.” U.S. Const. amend. XIV, § 2. Counting over 330 million people across 3.8 million square miles is a very difficult and complex task. Each decennial census takes over a decade to plan, execute, and complete, and involves myriad operational decisions. The 2020 decennial census—a 15.6-billion-dollar operation—is monitored and managed using a master schedule with over 27,000 separate lines of census activities, and is supported by no fewer than 52 separate information-technology systems.

The decennial census is also very important. It underpins our Nation’s representative democracy. It is used to allocate political power at all levels of government. And the data it collects and produces are used for countless purposes by governments, businesses, organizations, and individuals. Given the importance of the census, the Census Bureau must proceed carefully, with meticulous planning. Systems are developed, and tested, and tested again.

None of this would be possible without the cooperation of the public at large. Members of the public can be reluctant to reveal their and their household’s personal information to the government. But we ask them to do so every decade based on the promise—printed at the top of the census questionnaire—that their responses “are protected by law.”

This lawsuit concerns two large obstacles to the successful operation of the 2020 decennial census. The first obstacle is the COVID-19 pandemic, which unfortunately emerged just as hundreds of thousands of census field staff prepared to fan out around the country to collect information from the public. The once-in-a-century pandemic, along with major hurricanes and wildfires, caused a series of cascading delays that has rendered the Census Bureau unable to meet the statutory deadlines for delivering apportionment and redistricting data.

The second obstacle is the rise of computational power that threatens to reveal confidential information. It is now possible, using sophisticated algorithms on powerful systems, to reverse-engineer large sets of aggregated, supposedly de-identified data. Given this development, the Census Bureau set out to determine whether its data products were susceptible to such a “reconstruction attack.” And the Census Bureau determined—and third parties have confirmed—that the disclosure-avoidance method the Bureau applied to protect its 2010 data products no longer suffices to protect the confidentiality of census responses. If the Census Bureau were to continue doing what it did in 2010, it would be violating not only federal law, but also the confidentiality promise that it made to census respondents. And with that bond of trust broken, future census response rates would undoubtedly fall, and the accuracy of future censuses would suffer.

Plaintiffs—the State of Alabama, a congressional representative, and two individuals—would impose a third obstacle to the Census Bureau’s operations if the relief they seek through this lawsuit were granted. Plaintiffs first argue that the disclosure-avoidance method that the Census Bureau will apply to its forthcoming redistricting data products—differential privacy—will result in flawed numbers. They attempt to bolster their claim by relying on demonstration data that the Census Bureau specifically tuned to *amplify* the infusion of noise so that it could work with its data users to identify and mitigate issues in its various algorithms. But Plaintiffs acknowledge that the Census Bureau will release more-realistic demonstration data later this month. And, as Defendants explain below, those data—which will more-closely resemble the final redistricting data products—will be quite accurate. Plaintiffs nevertheless argue that *any* application of differential privacy will violate the Census Act on the grounds that the resulting data products would not constitute “tabulations of population.” But that argument is belied by the Census Act itself—as well as by Plaintiffs, who themselves refer to the Bureau’s forthcoming redistricting data products in their brief as tabulations of population.

The relief Plaintiffs seek also raises significant concerns. If this Court were to enjoin the use of differential privacy, the Bureau would still need to impose some form of disclosure avoidance. Plaintiffs suggest that the Bureau could use its ineffective 2010 disclosure-avoidance methodology for this year's census. But as explained below, any feasible alternative solution would result in far-less-accurate data and would take months to implement, at a minimum.

Though Plaintiffs ask that the Court prolong the extant delay, they also demand that Defendants produce the redistricting data now. But the redistricting data set does not yet exist, and will likely not come into existence in any form until late August, as the data are still being processed. To the extent that Defendants can produce the redistricting data earlier, they will do so. But any Order from this Court must take into account not only Plaintiffs' desires for the prompt publication of redistricting data, but also the reality that events beyond the Census Bureau's control have delayed the creation and production of those data products.

* * *

The decennial census is an extremely complicated endeavor. It is steered by expert scientists, statisticians, and systems engineers. It is the type of process that should be managed by subject-matter experts ultimately accountable to the elected Executive. "There is no basis for the judiciary to inject itself into this sensitive political controversy and seize for itself the decision to reevaluate the competing concerns between [census] accuracy and speed." *Nat'l Urban League v. Ross*, 977 F.3d 698, 713 (9th Cir. 2020) (Bumattay, J., dissenting from denial of administrative stay), *stay granted*, 141 S. Ct. 18 (2020). The same principle applies here: the Secretary of Commerce and the Census Bureau—not Plaintiffs or this Court—are best positioned to balance accuracy, confidentiality, and speed. Plaintiffs' motion and petition should be denied.

BACKGROUND

A. The Decennial Census

“The Constitution requires an ‘actual Enumeration’ of the population every 10 years and vests Congress with the authority to conduct that census ‘in such Manner as they shall by Law direct.’” *Wisconsin v. City of New York*, 517 U.S. 1, 5 (1996) (quoting U.S. Const. art. I, § 2, cl. 3). Congress, in turn, “has delegated to the Secretary of the Department of Commerce the responsibility to take ‘a decennial census of [the] population . . . in such form and content as he may determine.’” *Id.* (quoting 13 U.S.C. § 141(a)). “The Secretary is assisted in the performance of that responsibility by the Bureau of the Census and its head, the Director of the Census.” *Id.* (citing 13 U.S.C. §§ 2, 21).

“The Constitution provides that the results of the census shall be used to apportion the Members of the House of Representatives among the States.” *Id.* And “[b]ecause the Constitution provides that the number of Representatives apportioned to each State determines in part the allocation to each State of votes for the election of the President, the decennial census also affects the allocation of members of the electoral college.” *Id.* “[C]ensus data also have important consequences not delineated in the Constitution: The Federal Government considers census data in dispensing funds through federal programs to the States, and the States use the results in drawing intrastate political districts.” *Id.* at 5–6.

Today, the decennial census is a 15.6-billion-dollar operation, designed to count over 330 million people across 3.8 million square miles. *See* Declaration of Michael Thieme ¶¶ 4–5. And it necessarily requires the cooperation of the American public. For the 2020 census, the Census Bureau spent hundreds of millions of dollars to encourage the country to respond to the census, *see, e.g., id.* ¶ 12, and hundreds of thousands census field staff fanned out across the country to follow up on nonresponding addresses, *see id.* ¶¶ 4, 19–28.

“Although each [decennial census] was designed with the goal of accomplishing an ‘actual Enumeration’ of the population, no census is recognized as having been wholly successful in achieving that goal.” *Wisconsin*, 517 U.S. at 6. As a massive, human-driven operation, the census is, almost by definition, imperfect, despite the monumental efforts of the Census Bureau staff who strive to “count everyone living in the country once, only once, and in the right place.” Thieme Decl. ¶ 3. “Persons who should have been counted are not counted at all or are counted at the wrong location; persons who should not have been counted (whether because they died before or were born after the decennial census date, because they were not a resident of the country, or because they did not exist) are counted; and persons who should have been counted only once are counted twice.” *Wisconsin*, 517 U.S. at 6. As a result, census data “may be as accurate as such immense undertakings can be, but they are inherently less than absolutely accurate.” *Gaffney v. Cummings*, 412 U.S. 735, 745 (1973).

B. The Census Act’s Confidentiality Provisions

“[A]n accurate census,” of course, “depends in large part on public cooperation.” *Baldrige v. Shapiro*, 455 U.S. 345, 354 (1982). But many people chafe at the notion of providing the government with their personal information. Census Bureau research shows that over half of census respondents were at least “somewhat concerned” – with 28% “very concerned” or “extremely concerned” – about the confidentiality of their census responses. Declaration of John M. Abowd ¶ 11. And “[t]hese concerns are even more pronounced in minority populations and represent a major operational challenge to enumerating traditionally hard-to-count populations.” *Id.*

“To stimulate [the public’s] cooperation Congress has provided assurances that information furnished to the Secretary by individuals is to be treated as confidential.” *Baldrige*, 455 U.S. at 354 (citing 13 U.S.C. §§ 8(b), 9(a)). In particular, sections 8 and 9 of the Census Act provide in part that: (i) “the Secretary [of Commerce] may furnish copies

of tabulations and other statistical materials which do *not* disclose the information reported by, or on behalf of, any particular respondent,” 13 U.S.C. § 8(b) (emphasis added); and (ii) Defendants, and their officers and employees, may not “make *any* publication whereby the data furnished by any particular establishment or individual under this title can be identified,” 13 U.S.C. §§ 9(a), (a)(2) (emphasis added). Indeed, the Census Act provides that Census Bureau staff that publish information protected by § 9 “shall be” subject to fines “or imprisoned not more than 5 years, or both.” 13 U.S.C. § 214. In short, “§ 8(b) and § 9(a) of the Census Act embody explicit congressional intent to preclude *all* disclosure of raw census data reported by or on behalf of individuals.” *Baldrige*, 455 U.S. at 361 (emphasis added).

C. The Rise of Computing Power and Its Implications for Confidentiality

In past decennial censuses, the Census Bureau protected the confidentiality of the released data by using disclosure-avoidance mechanisms such as suppression (*i.e.*, withholding data) and, in later censuses, data-swapping (*i.e.*, where certain characteristics of a number of households are swapped with those of other households as paired by a matching algorithm). Abowd Decl. ¶¶ 23–25. The 2010 decennial census employed data-swapping as its primary disclosure-avoidance mechanism, and the Census Bureau’s data-swapping methodology kept the total population and total-voting-age population constant for each census block, the smallest level of census geography. *Id.* ¶ 25. This method of disclosure avoidance was considered sufficient at the time. *See id.* ¶¶ 26, 49.

That is no longer the case. It has long been known that purportedly de-identified, aggregated data may be “reconstructed” by a series of mathematical algorithms, though such attacks had been constrained by the limits of available computational power. In one famous example, Professor Latanya Sweeney revealed in 1997 that she had re-identified then-Massachusetts Governor William Weld’s medical records in a purportedly de-identified public database. *See id.* ¶ 27. And as computing power becomes cheaper, more

plentiful, and more accessible as it moves to the cloud, re-identification attacks have increased, and have targeted increasingly large datasets. One recent article recounted re-identification attacks on supposedly de-identified datasets as varied as German internet browsing histories, Australian medical records, New York City taxi trajectories, and London bike-sharing trips. *See* Luc Rocher *et al.*, “Estimating the success of re-identifications in incomplete datasets using generative models,” *Nature Communications* (2019), available [here](#); *see also* Abowd Decl. ¶¶ 33–36 (collecting other examples).

The decennial census is not immune to these trends. Following the 2010 census, the Census Bureau published *over 150 billion independent statistics* about the characteristics of the 308,745,538 persons enumerated in the census. Abowd Decl. ¶ 18. The Census Bureau thus conducted its own reconstruction experiment based on just 6.2 billion of those statistics. The Bureau’s simulated attack precisely reconstructed approximately 46% of the 308,745,538 records with their exact race, ethnicity, sex, *and* age—and more than 70% of the reconstructed records had exact race, ethnicity, and sex, and were within one year of actual age. *See* Abowd Decl. App’x B ¶¶ 5–7.

The Census Bureau then attempted a re-identification experiment using commercially available databases, and was able to successfully re-identify about 52 million individuals—roughly 17% of the people enumerated in the 2010 census. *See id.* ¶¶ 22–23; Abowd Decl. ¶ 38. And if an attacker had access to data better than the third-party data used in the Census Bureau’s simulation, as many as 179 million people could correctly be re-identified. *See* Abowd Decl. App’x B ¶¶ 24; Abowd Decl. ¶ 38. Although Dr. Abowd had in 2018 described the re-identification risk as “small,” he retracted that tentative conclusion at the February 16, 2019, session of the American Association for the Advancement of Science. *See* Abowd Decl. ¶ 83.

This serious reconstruction and re-identification vulnerability has been confirmed by the JASON group, which Plaintiffs describe as “an independent group of scientists and engineers from whom the Census Bureau has sought third-party review,” and on

whose work Plaintiffs rely. Pls. Mot., Doc. 3 (“Mot.”) at 31. The JASON group explained—in a publication that Plaintiffs repeatedly cited to the Court, *see* Mot. 13 & n.24, 29 & n.57, 31, 32 & nn.58–59—that, in its view, “Census has convincingly demonstrated the existence of a vulnerability that census respondents can be re-identified through the process of reconstructing microdata from the decennial census tabular data and linking that data to databases containing similar information that can identify the respondent.” *See generally* JASON, *Formal Privacy Methods for the 2020 Census* (Apr. 2020) at 89, available [here](#). The JASON group summarized its findings on this point as:

- The Census has demonstrated the re-identification of individuals using the published 2010 census tables.
- Approaches to disclosure avoidance such as swapping and top and bottom coding applied at the level used in the 2010 census are insufficient to prevent re-identification given the ability to perform database reconstruction and the availability of external data.

Id. at 6; *accord id.* at 93–94. In short, as Dr. Abowd explains, data produced by the 2010 disclosure-avoidance mechanism would be “vulnerable to reconstruction and re-identification attacks because of the parameters of the swapping mechanism’s 2010 implementation: an overall insufficient level of noise, the invariants preserved without noise, and the geographic and demographic detail of the published summary data.” Abowd Decl. ¶ 39. As such, “[t]he Census Bureau can no longer rely on the swapping implementation used in 2010 if it is to meet its obligations to protect respondent confidentiality.” *Id.*; *see generally id.* ¶¶ 41–43, 50–51.

D. Differential Privacy

At a fundamental level, all disclosure-avoidance methodologies have a necessary impact on the availability and accuracy of the resulting data. That is how confidentiality is protected. Data-swapping, for example, injects noise into the census redistricting data by swapping certain characteristics between a subset of households. *See* Abowd Decl.

¶¶ 25. But data-swapping – as demonstrated by the Census Bureau and corroborated by the JASON group – is susceptible to database reconstruction attacks. *See id.* ¶¶ 26, 39. And the precise data-swapping methodology used is necessarily opaque, so as to better protect the confidentiality of the data. As Dr. Abowd explains, “[i]mplementation parameters for these legacy disclosure avoidance methods, especially swapping rates, are often some of the most tightly guarded secrets that the Census Bureau protects.” Abowd Decl. ¶ 62.

Given the now-demonstrable flaws with the disclosure-avoidance methodologies used in the 2010 decennial census, “a swapping mechanism that targets vulnerable households for swapping would require significantly higher rates of swapping than were used in 2010 to protect against a reconstruction attack.” *Id.* ¶ 42. And utilizing such higher swapping rates would “have a significant, detrimental impact on data quality.” *Id.* Moreover, “[i]mplementing swapping in 2020 would also require abandoning the total population and voting-age population invariants that were used in 2010” for two reasons: (i) it would be “impossible to find enough paired households with the same number of persons and adults without searching well outside the neighborhood of the original household”; and (ii) “holding the total and adult populations invariant gives the attacker a huge reconstruction advantage—exact record counts in each block for persons and adults”—and that advantage “vastly improves the accuracy of the reconstructed data.” *Id.* But “[i]nternal experiments . . . confirmed that increasing the swap rate from the level used in 2010 and removing the invariants on block-level population counts (to permit the increased level of swapping and protect against reconstruction attacks) would render the resulting data unusable for most data users.” *Id.*

Nor is data suppression a viable option. “While the Census Bureau could use suppression to protect from a reconstruction attack, the resulting data would be only available at a very high level of generality.” *Id.* ¶ 43. “Today’s data users, including redistricters, rely on detailed block and tract-level data, which would not be available for

many areas if the Census were to return to suppression to protect against modern attacks.” *Id.*

Ultimately, the Census Bureau’s Data Stewardship Executive Policy Committee (DSEP) determined that neither swapping nor suppression would allow the Census Bureau “to produce high quality statistics from the decennial census while also protecting the confidentiality of respondents’ census records” as required by the Census Act. *Id.* ¶ 46; *see also id.* ¶ 51 (“[T]o achieve the necessary level of privacy protection, both enhanced data swapping and suppression had severely deleterious effects on data quality and availability.”).

This led the Census Bureau to differential privacy, “[t]he best disclosure avoidance option that offers a solution capable of addressing the new risks of reconstruction-abetted re-identification attacks, while preserving the fitness-for-use of the resulting data for the important governmental and societal uses of census data.” *Id.* ¶ 47. Differential privacy is used by major private-sector technology firms, and the Census Bureau has been using differential privacy to protect certain of its statistical products since 2008. *See id.* ¶ 45.

“Differential privacy, first developed in 2006, is a framework for quantifying the precise disclosure risk associated with each incremental release from a confidential data source.” *Id.* ¶ 44. This framework allows “the Census Bureau to quantify the precise amount of statistical noise required to protect privacy.” *Id.* “This precision allows the Census [Bureau] to calibrate and allocate precise amounts of statistical noise in a way that protects privacy while maintaining the overall statistical validity of the data.” *Id.* The amount of noise injected is determined by a measure known as the privacy-loss budget (PLB) or the “epsilon.” Michael Hawes, U.S. Census Bureau, “Differential Privacy and the 2020 Decennial Census” (Mar. 5, 2020), at 18, available [here](#). Setting epsilon to zero would result in perfect privacy but useless data, and setting the epsilon to infinity would result in perfect accuracy, but would result in releasing data in fully identifiable form. *Id.*

The advantages of differential privacy are myriad. *See, e.g.,* Simson L. Garfinkel, U.S. Census Bureau, *Modernizing Disclosure Avoidance* (Sept. 15, 2017) at 10, available [here](#). Those advantages include protection against database reconstruction attacks and privacy guarantees that do not depend on the availability of external data. *See id.* It can do so while still producing highly accurate data. Abowd Decl. ¶ 54. And, as will be implemented by the Census Bureau, the accuracy of the data increases, not decreases, as census geographies increase in size. *See id.* ¶ 56.

Moreover, differential privacy can be tuned to determine the optimal setting whereby the privacy of confidential data can be reasonably assured, yet the resulting data will be fit for redistricting and other uses. *See id.* ¶¶ 52, 54, 59. The Bureau’s “empirical analysis showed that differential privacy offered the most efficient trade-off between privacy and accuracy — [its] calculations showed that the efficiency of differential privacy dominated traditional methods.” *Id.* ¶ 41. “In other words, regardless of the level of desired confidentiality, differential privacy will always produce more accurate data than the alternative traditional methods considered by the Census Bureau.” *Id.*

Differential privacy also allows for unprecedented transparency. “The Census Bureau has submitted its differential privacy mechanisms, programming code, and system architecture to thorough outside peer review.” Abowd Decl. ¶ 62. The Bureau has “also committed to publicly releasing the entire production code base and full suite of implementation settings and parameters.” *Id.* Whereas swapping techniques “must be implemented in a ‘black box,’” to protect the resulting data, differential privacy, by contrast, “does not rely on the obfuscation of its implementation as a means of protecting the data.” *Id.* “The Census Bureau’s transparency will allow any interested party to review exactly how the algorithm was applied to the 2020 Census data, and to independently verify that there was no improper or partisan manipulation of the data.” *Id.*

And the Census Bureau has aimed to tune the disclosure-avoidance algorithms, and will tune the privacy-loss budget, in the public eye. *See generally id.* ¶¶ 57–62. In

October 2019 and throughout 2020, the Census Bureau publicly released “demonstration data.” See U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). Exactly as designed, these public releases resulted in “extensive actionable feedback from the data user community,” which “has informed ongoing [disclosure-avoidance] system improvements and design changes.” *Id.* During this iterative process, the Census Bureau “used a lower privacy-loss budget than [it] anticipate[s] using for the final 2020 Census data—that is, these demonstration data were purposefully ‘tuned’ to privacy and not ‘tuned’ for producing highly accurate redistricting data.” Abowd Decl. ¶ 61. The Bureau did so in order “to home in on the elements of the algorithm that were causing systemic distortions that needed to be addressed.” U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). This decision “meant that the resulting [demonstration] data would have substantially more noise (error) than should be expected in the final 2020 Census data products,” but it “unfortunately led some of our data users to expect comparable amounts of noise in the final 2020 Census data.” *Id.*

Fortunately, that will not be the case. By keeping the privacy-loss budget roughly constant in the demonstration data to date, the Census Bureau has been able to improve the post-processing algorithms and mitigate post-processing errors. See U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 3, 2021), available [here](#).¹ For example, “the Census Bureau has identified and corrected the algorithmic sources of [certain] distortions,” and “any residual impact of the types of systematic bias observed in the early

¹ The *amicus* States prove this point. They note, for example, that Utah “analyzed the 2010 demonstration data, comparing it with the previously received 2010 redistricting data and sent its findings to the Census Bureau.” Doc. 40 at 2. And they acknowledge that this iterative process worked: Utah acknowledges that it “saw an improvement from the October 2019 to the November 2020 demonstration data,” *id.* at 3, though they incorrectly attribute that improvement to modifications in the privacy-loss budget. See *id.*

demonstration data will be negligible and well within the normal variance and total error typical for a census.” Abowd Decl. ¶ 67.

And with those algorithmic improvements in place, the Census Bureau moved to tuning the privacy-loss budget. “On March 25, 2021, DSEP approved the privacy-loss budget to be used for the next demonstration product. This privacy-loss budget reflects empirical analysis of over 600 full-scale runs of the Disclosure Avoidance System using 2010 Census data.” Abowd Decl. ¶ 70. “The Census [Bureau] evaluated these experimental runs using accuracy and fitness-for-use criteria for the redistricting use case informed by the extensive feedback we have received from the redistricting community and the Civil Rights Division at the U.S. Department of Justice.” *Id.*

The Census Bureau intends to release the next set of demonstration data by April 30, 2021. *See* U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). This set of data employs a higher privacy-loss budget, tuned for accuracy, “that better approximates the final privacy-loss budget that will likely be selected for the redistricting data product.” Abowd Decl. ¶ 69. “These new demonstration data will also reflect system design changes that have been made since the last demonstration data release, along with tuning and optimization of the system that have been done specifically to prioritize population count accuracy and the ability to identify majority-minority districts.” *Id.*

“The next iteration of demonstration data will establish that differential privacy protections can produce extremely accurate redistricting data.” Abowd Decl. ¶ 54. In the upcoming release of demonstration data:

- “Total populations for counties have an average error of +/- 5 persons . . . as noise from differential privacy” (an error rate of about 0.04% of the counties’ population). Compare that level of precision with the “average county-level” estimated uncertainty inherent in census counts, which “is +/- 960 persons (averaging 1.6% of the county census counts).” *Id.*

- “At the block level the differentially private data have an average population error of +/- 3 persons” which is also more precise than “the simulated error inherent in the census which puts the average error uncertainty of block population counts at +/- 6 people.” *Id.*
- “In the April 2021 Demonstration Data Product, Congressional districts as drawn in 2010 [nationwide] have a mean absolute percentage error of 0.06%.” *Id.* ¶ 56.
- “Even for state legislative districts, which had average sizes of 159,000 (upper chambers) and 64,000 (lower chamber[s]), the mean absolute percentage errors are 0.09% (upper chambers) and 0.16% (lower chambers), respectively. Such errors are trivial and imply that the difference between districts drawn from the April 2021 Demonstration Data Product and those drawn from the original 2010 P.L. 94-171 Redistricting Data Summary File would be statistically and practically imperceptible.” *Id.*
- “The April 2021 demonstration data show no meaningful bias in the statistics for racial and ethnic minorities even in very small population geographies like Federal American Indian Reservations.” *Id.* ¶ 55 (emphasis omitted). “The data permit assessment of the largest OMB-designated race and ethnicity group in each geography – the classification used by the Department of Justice for Voting Rights Act scrutiny – with a precision of 99.5% confidence in variations of +/- 5 percentage points for off-spine geographies as small as 500 persons, approximately the minimum voting district size in the redistricting plans that the Department of Justice provided as examples.” *Id.*

In sum, the demonstration data that will be released later this month will demonstrate that the differential-privacy algorithm, “when properly tuned, ensures that redistricters

can remain confident in the accuracy of the population counts and demographic characteristics of the voting districts they draw, despite the noise in the individual building blocks.” *Id.* ¶ 56 (emphasis omitted).

Data-users will have at least four weeks to review the next set of demonstration data, perform their analyses, and submit feedback. *See* U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). In early June, DSEP will set the final privacy-loss budget and production parameters for the redistricting data product. *See id.* Applying differential privacy to the redistricting data will take roughly three weeks—“similar to the period required to implement disclosure avoidance in prior censuses”—and “is not the cause of the delay in the delivery of the redistricting data.” Abowd Decl. ¶ 72. In fact, “the disclosure avoidance procedures completed in the 2010 census processing took 27 days--or nearly *four* weeks.” Thieme Decl. ¶ 71 (emphasis added).

To the contrary, shifting disclosure-avoidance methodologies now is all but guaranteed to cause further delay—and “[t]he effect on the schedule for delivering redistricting data would be substantial.” Abowd Decl. ¶¶ 84–85. “[U]nder all scenarios the delay would be *multiple months*.” *Id.* ¶ 85 (emphasis added). “This delay is unavoidable because the Census Bureau would need to develop and test new systems and software, then use them in production and subject the results to expert subject matter review prior to production of data.” *Id.*

Because the 2010 census data are vulnerable to a database reconstruction attack, “the Census Bureau cannot simply repeat the swapping protocols from the 2010 census, but rather would be forced to fashion appropriate levels of protection”—and “[u]sing an appropriate level of protection for either suppression or swapping would produce far less accurate data than would differential privacy.” *Id.* ¶ 87. And even if the Census Bureau were “ordered to repeat exactly what was done in 2010 (despite the serious risks to privacy the Census has identified),” the Bureau “could not simply ‘flip a switch’ and

revert to the prior methodology.” *Id.* ¶ 86. “The 2020 Census’s system architecture is completely different than that used in the 2010 Census, and it is thus not possible to simply ‘plug in’ the disclosure-avoidance system used in 2010.” *Id.* “Instead,” the Bureau “would need to conduct the requisite software development and testing.” *Id.*

Simply put, it is not practical at this late hour to change the disclosure-avoidance system’s methodologies. Such decisions “are highly technical and can have unanticipated consequences.” *Id.* ¶ 88. “While [the Census Bureau] cannot predict the full impact of any change, there is a danger than any change would have cascading effects on data accuracy and privacy, making race and ethnicity data, along with age data, substantially less accurate.” *Id.* And “[a]ny sort of change in the basic methodology would be minimally tested and would not have the benefit of any input from the user community.” *Id.*

E. The Census Bureau’s Delivery of Redistricting Data

As explained above, the 2020 Census has been a massive undertaking. While the Bureau has done everything in its power to complete the census as expeditiously as possible, the COVID-19 pandemic has resulted in some unavoidable delay. The original plan was for the Census Bureau to begin in-person operations (called Nonresponse Followup or NRFU) in May 2020, but it was forced to suspend those operations for months due to the pandemic. Thieme Decl. ¶ 30. By the time the Census Bureau entered the field in earnest three months later, it did so during a perfect storm of natural disasters and civil unrest. *Id.* ¶ 33. “Devastating hurricanes in the Gulf Coast area . . . limited and slowed the Census Bureau’s ability to conduct NRFU operations.” *Id.* In “large areas of the West Coast, field operations were hampered by conflagrations that caused health alerts due to fire and smoke.” *Id.* And “in cities across the country,” civil unrest made the already-difficult enumeration even harder. *Id.*

Making matters worse, the Secretary and the Census Bureau were under a statutory directive to report the census results to the President by December 31, 2020 so that he could timely submit them to Congress for reapportionment of the House. *See* 13 U.S.C.

§ 141(b); 2 U.S.C. § 2a. And although the Secretary had asked for an extension of these statutory deadlines, Congress did not oblige. Thieme Decl. ¶ 35. So the Census Bureau again adjusted its operations in an attempt to meet the statutory deadlines. *Id.* ¶ 36. But that adjustment led to the intervention of another Branch: the Judiciary. After a court-ordered preliminary injunction forced the Census Bureau to remain in the field, an emergency Supreme Court ruling stayed that injunction and allowed the Census Bureau to conclude field operations in mid-October 2020, having resolved 99.9% of all housing units in the process. *See Ross v. Nat'l Urban League*, 141 S. Ct. 18 (2020); Thieme Decl. ¶ 36.

But collecting responses through completed questionnaires and in-person field work is not the end of the story – the Census Bureau must then summarize the individual and household data that it collected into usable, high-quality tabulations. Thieme Decl. ¶¶ 37–83. Although creating such tabulations may appear easy, it is not. The Census Bureau must integrate data from different enumeration methods used across the country, identify any issues or inconsistencies that arise, rectify them, and produce tabulations that will guide the country for the next ten years, all without compromising its statutory mandate to maintain the confidentiality of census responses. 13 U.S.C. §§ 8, 9; Thieme Decl. ¶¶ 53–59 (describing how administrative records are incorporated and data are reconciled to produce the Census Unedited File); *id.* ¶¶ 60–64 (describing how the federally affiliated overseas population is incorporated into the data to produce apportionment numbers); *id.* ¶¶ 65–70 (describing the iterative process for compiling detailed information such as race, ethnicity, and age to produce the Census Edited File); *id.* ¶¶ 71–74 (describing the process for applying the Census Bureau’s disclosure-avoidance methodology); *id.* ¶¶ 75–78 (describing the process for generating usable data files).

Even working with all possible dispatch, the Census Bureau was not able to meet its December 31, 2020 statutory deadline for reporting apportionment numbers. Due to the difficulties encountered during data collection and issues that arose during the processing phase, the Census Bureau projects that it will not complete apportionment counts

until April 30, 2021. Thieme Decl. ¶ 37. Another court and other parties have even relied upon Defendants' representation that "the Census Bureau will not under any circumstances report the results of the 2020 Census . . . before April 16, 2021." *Nat'l Urban League v. Raimondo*, No. 20-cv-05799, ECF Nos. 465 & 467 (N.D. Cal. Feb. 3, 2021).

The delay in producing apportionment data also means the Secretary and the Census Bureau have missed the statutory deadline (March 31, 2021) to submit census-based redistricting data to the States. 13 U.S.C. § 141(c). This was not a secret. In a February 12, 2021 Press Release, the Census Bureau explained that "it will deliver the [] redistricting data to all states by Sept. 30, 2021" because "COVID-19-related delays and prioritizing the delivery of the apportionment results delayed the Census Bureau's original plan to deliver the redistricting data to the states by March 31, 2021." *Census Bureau Statement on Redistricting Data Timeline*, U.S. Census Bureau (Feb. 12, 2021), available [here](#).

That announcement was not for the Census Bureau's benefit, but for States that use census-based redistricting data to draw their congressional or state election districts. While no federal law requires the use of census data for this purpose, the data are generally utilized as the gold standard, including by the Department of Justice, which uses such data for enforcement of the Voting Rights Act. Declaration of James Whitehorne ¶ 4. That's why States generally use census data for redistricting. And many of those States make up the 27 States that are bound by their own laws to redistrict in 2021. *See 2020 Census Delays and the Impact on Redistricting*, National Conference of State Legislatures (last visited Apr. 11, 2021), available [here](#). That has led some States under self-imposed redistricting pressure to find workable solutions. In New Jersey, for example, voters approved a constitutional amendment that allowed the State to use previous district maps until the new maps are in effect for the 2023 elections. *See Whitehorne Decl.* ¶ 7; N.J. Const. art. IV, § 3, ¶ 4. And in California, the state legislature sought and obtained at least a four-month delay of the redistricting deadlines from the California Supreme Court. *Legislature of the State of Cal. v. Padilla*, 469 P.3d 405, 413 (Cal. 2020);

Whitehorne Decl. ¶ 7. These States – and many others – gathered information from the Census Bureau and found a way to remedy their own redistricting issues. Whitehorne Decl. ¶¶ 7–8.

Alabama is not one of those States. Instead, Alabama now seeks redistricting data that does not exist by a statutory deadline that is impossible to meet. Whitehorne Decl. ¶¶ 14–16. Defendants oppose that request.

ARGUMENT

I. PLAINTIFFS LACK STANDING.

“The doctrine of standing is an essential and unchanging part of the case-or-controversy requirement embodied in Article III of the Constitution.” *Flat Creek Transp., LLC v. Fed. Motor Carrier Safety Admin.*, 923 F.3d 1295, 1300 (11th Cir. 2019).² “In the absence of standing, a court is not free to opine in an advisory capacity about the merits of a plaintiff’s claims, and the court is powerless to continue.” *Aaron Private Clinic Mgmt. LLC v. Berry*, 912 F.3d 1330, 1335 (11th Cir. 2019).

“The irreducible constitutional minimum of standing requires a plaintiff to show that he (1) suffered an injury in fact, (2) that is fairly traceable to the challenged conduct of the defendant, and (3) that is likely to be redressed by a favorable judicial decision. *Flat Creek Transp., LLC*, 923 F.3d at 1300. “[A]s the part[ies] invoking federal jurisdiction,” Plaintiffs “bear[] the burden of establishing these elements.” *Id.* “And because standing doctrine is intended to confine the federal courts to a properly judicial role,” those courts must “take seriously the requirement that a plaintiff *clearly* demonstrate each requirement.” *Id.* (emphasis added). “If the plaintiff fails to meet its burden, this court lacks the power to create jurisdiction by embellishing a deficient allegation of injury.” *Aaron Private Clinic Mgmt. LLC*, 912 F.3d at 1336.

² Unless expressly included, all citations and internal quotation and alteration marks have been omitted.

Plaintiffs have not demonstrated—let alone “clearly” demonstrated—any of the three necessary standing elements. Accordingly, their motion should be denied.

A. Plaintiffs Have Not Sustained Any Injuries-in-Fact

An “injury in fact” is “the invasion of a judicially cognizable interest that is concrete and particularized and actual and imminent.” *Corbett v. Transp. Sec. Admin.*, 930 F.3d 1225, 1228 (11th Cir. 2019). Plaintiffs have not demonstrated that any of them have been injured or will imminently be injured, either by the application of differential privacy, or by the delay in producing the redistricting data.

1. Plaintiffs Are Not Injured by Differential Privacy

Plaintiffs assert five forms of injury-in-fact in connection with their differential-privacy claims. None has merit.

a. Informational Injury

Asserting a supposed informational injury, Plaintiffs argue that Alabama is statutorily entitled to “tabulations of population” under 13 U.S.C. § 141(c), *see* Mot. 29–33; Compl. ¶¶ 133–140—and that is precisely what the Secretary will provide to the State. Plaintiffs acknowledge that the term “‘tabulate’ has long been understood to mean ‘[t]o put or arrange in a tabular, systemic, or condensed form.’” Mot. 29 n.57 (quoting *The Random House College Dictionary* 1337 (revised ed. 1975)). It follows that a “tabulation” is the arrangement of data in such form. And Plaintiffs do not dispute that the Secretary will provide to the State data in such an arranged form. Hence, Alabama will receive “tabulations.”

One need only review Plaintiffs’ brief to confirm this fact. Plaintiffs contend that the “*tabulations*” “will be intentionally scrambled.” *Id.* at 2. They allege that they will suffer harm from supposedly “flawed *tabulations*.” *Id.* at 4 (emphasis added). They express concern about supposedly “false *tabulations*.” *Id.* at 27 (emphasis added). They argue that “Defendants plan to provide the State with inaccurate *tabulations*.” *Id.* at 34 (emphasis added). And they contemplate what might happen if “both *tabulations*”—*i.e.*,

tabulations with and without the application of differential privacy — “can be released.” *Id.* at 55 (emphasis added). Plaintiffs may not agree with the methodology that will underlie the Secretary’s tabulations, but Plaintiffs readily acknowledge that they are, in fact, tabulations.

These tabulations will further constitute the “tabulations of population” contemplated in § 141(c). Plaintiffs do not contend that the Secretary will simply invent population numbers. Rather, to ensure compliance with the confidentiality requirements imposed by Congress, *see* 13 U.S.C. §§ 8 & 9, the Census Bureau will inject slight statistical “noise” into the sub-state population counts. *See, e.g.,* Abowd Decl. ¶¶ 54, 69. But that process hardly renders the resulting data something other than “tabulations of population.”

Again, Plaintiffs themselves prove the point. They claim that the Secretary will, in their view, “provide the States purposefully flawed *population tabulations*.” Mot. 1–2 (emphasis added). They contend that “[i]f the Census Bureau uses differential privacy, the *population tabulations* it reports to States for redistricting will be inaccurate.” *Id.* at 24 (emphasis added); *accord id.* at 25. They represent that “[t]he Court will be unable to remedy” supposed “harms if Defendants deliver *population tabulations* infected by differential privacy.” *Id.* at 27 (emphasis added). They argue about what might happen “once the skewed *population tabulations* are delivered.” *Id.* at 51 (emphasis added). And they talk about losing funding “if the *population tabulations* are inaccurate.” *Id.* at 52 (emphasis added); *see also, e.g., id.* at 4 (characterizing differential privacy as “a ‘statistical method’” used “‘to determine the population for purposes of . . . redistricting’”); Pls. Reply, Doc. 25, at 4 (“Challenges to statistical methods that ‘determine the population for purposes of the apportionment or redistricting’ must be heard by a three-judge court.”) (emphasis omitted). But they admit that the tabulations that the Secretary will deliver are, in fact, “tabulations of population.”

In an effort to call into question future population tabulations, Plaintiffs point to their experts' analysis of the Census Bureau's releases of demonstration data. *See generally* Mot. 18–24. Yet Plaintiffs acknowledge that “[f]or the demonstration data products, the Census Bureau set a more conservative privacy-loss budget than it expects will be set for the 2020 census—meaning that the demonstration data will have more ‘noise (error)’ than should be expected in the final 2020 Census data products.” *Id.* at 18 (quoting U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021)).

In fact, the Census Bureau explained that it maintained this conservative privacy-loss budget—even though doing so “meant that the resulting data would have *substantially* more noise (error) than should be expected in the final 2020 Census data products”—so the Bureau and its data users could “home in on the elements of the algorithm that were causing systemic distortions that needed to be addressed.” U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#) (emphasis added). The Census Bureau is planning to release the next set of demonstration data on April 30, 2021. *Id.*; *see* Mot. 49 (acknowledging same). That demonstration data: (i) “will feature a higher [privacy-loss budget] and system parameter optimization informed by the hundreds of full-scale [disclosure-avoidance system] experimental runs [the Bureau has] been performing over the last several months”; (ii) “will more closely approximate the expected accuracy and fitness-for-use of the final 2020 Census redistricting data product”; and (iii) “will enable [the Bureau’s] data users to provide critical fitness-for-use analyses” and to “submit feedback and recommendations prior to” the Bureau’s Data Stewardship Executive Policy Committee’s decision that will set the final privacy-loss budget in June. U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). Indeed, the average population error in the forthcoming April 30 demonstration data falls well within the estimated uncertainty inherent in the census. *See* Abowd Decl. ¶ 54; *see supra* Background Part D.

Because Plaintiffs do not know how the privacy-loss budget will ultimately be set, or how that future budget will affect the redistricting data, their challenge to differential privacy is facial in nature. Plaintiffs concede as much. Admitting that the final redistricting data will be subject to less noise than the demonstration data to date, Plaintiffs argue that “no matter where the epsilon value is set,” the redistricting data “will just be less wrong than the demonstration numbers were,” and that “*any* application of differential privacy will produce erroneous numbers.” Mot. 18, 35 (emphasis added). In other words, Plaintiffs acknowledge their burden on this facial challenge: they “must establish that no set of circumstances exists under which” the application of differential privacy “would be valid.” *Reno v. Flores*, 507 U.S. 292, 301 (1993) (no-set-of-circumstances test applies to “both the constitutional challenges . . . and the statutory challenge”); *accord*, e.g., *Associated Builders & Contractors of Tex., Inc. v. Nat’l Labor Relations Bd.*, 826 F.3d 215, 220 (5th Cir. 2016); *Scherer v. U.S. Forest Serv.*, 653 F.3d 1241, 1243 (10th Cir. 2011) (Gorsuch, J.); *Sherley v. Sebelius*, 644 F.3d 388, 397 (D.C. Cir. 2011). And “[t]his heavy burden makes such an attack the most difficult challenge to mount successfully.” *Doe v. Kearney*, 329 F.3d 1286, 1294 (11th Cir. 2003).

Plaintiffs’ effort to satisfy their heavy burden rests on the theory that the “tabulation of total population by States” referenced in § 141(b) is equivalent to the “actual population counts for States,” and “[i]t follows that the ‘tabulations of population’ referenced in subsection 141(c) must also be the actual population counts.” Mot. 30. But nothing in § 141(b) suggests that the term “tabulation” contemplates any particular methodology. The methodology used to determine the apportionment counts stems from the *Constitution*, which requires that the apportionment of Representatives be based on an “actual Enumeration.” U.S. Const. art. I, § 2, cl. 3; *see Dep’t of Commerce v. U.S. House of Representatives*, 525 U.S. 316, 346–47 (1999) (Scalia, J., concurring) (“Dictionaries roughly contemporaneous with the ratification of the Constitution demonstrate that an ‘enumeration’

requires an actual counting, and not just an estimation of number.”). Section 141(b) references only “[t]he *tabulation* of total population by States,” 13 U.S.C. § 141(b) (emphasis added), and not, for example, “[t]he *enumeration* of total population by States.” It does not make sense, then, for Plaintiffs to attempt to synonymize “tabulation” with “enumeration.” Cf. *Firststar Bank, N.A. v. Faul*, 253 F.3d 982, 991 (7th Cir. 2001) (noting “the canon that different words within the same statute should, if possible, be given different meanings”). Instead, Congress used the term “tabulation of total population” in § 141(b) to mean exactly what it says—and how Plaintiffs use it repeatedly in their brief, *see supra*: an arrangement of population data for transmission to the President. *Conn. Nat’l Bank v. Germain*, 503 U.S. 249, 253–54 (1992) (“We have stated time and again that courts must presume that a legislature says in a statute what it means and means in a statute what it says there.”). Put simply, Plaintiffs’ invocation of the obvious—that the word “tabulation” appears in both § 141(b) and § 141(c)—is a non sequitur; it proves only that Congress wanted the Secretary to arrange population data for two different distributions.

And even if the term “tabulation” in § 141(b) could be construed to incorporate a particular methodology, the Census Act itself disproves the notion, *contra* Mot. 30, that any such methodology carries over to § 141(c). For example, the data that underlie the § 141(c) tabulations may be based on statistical sampling, whereas the data that underlie the § 141(b) tabulation may not. Section 195 of the Census Act provides that “the Secretary shall, if [s]he considers it feasible, authorize the use of the statistical method known as ‘sampling’ in carrying out the provisions of this title” — “[e]xcept for the determination of population for purposes of apportionment of Representatives.” 13 U.S.C. § 195. So the data that underlie “[t]he tabulation of total population by States . . . as required for the apportionment of Representatives,” § 141(b), cannot be premised on statistical sampling. But § 195 expressly provides that determinations of population for non-apportionment purposes—such as the redistricting data contemplated by § 141(c)—may properly be based on statistical sampling. *See, e.g., Glavin v. Clinton*, 19 F. Supp. 2d 543, 552–53 (E.D.

Va. 1998) (three-judge court) (“[T]he only plausible interpretation of the plain language and structure of the Act is that Section 195 prohibits sampling for apportionment and Section 141 allows it for all other purposes.”), *aff’d sub nom., Dep’t of Commerce v. U.S. House of Representatives*, 525 U.S. 316 (1999). In other words, nothing in the Census Act would preclude the Secretary from both: (i) producing the “tabulation of total population by States . . . as required for the apportionment of Representatives” under § 141(b) based on the actual enumeration; and (ii) developing the sub-state “[t]abulations of population” contemplated by § 141(c) through, say, a hybrid enumeration-and-statistical-sampling protocol.

This point is further borne out by the drafting history of the Census Act. Congress added § 141(c) in December 1975 but did not at that time amend § 195 to carve out the § 141(c) tabulations from § 195’s statistical-sampling authorization. *See* Pub. L. No. 94–171, 89 Stat. 1023 (Dec. 23, 1975). And less than a year later, Congress amended both § 141(c) and § 195. *See* Pub. L. No. 94–521 §§ 7(a) & 10, 90 Stat. 2459 (Oct. 17, 1976). But Congress again declined to carve out the § 141(c) tabulations from § 195’s statistical-sampling authorization. Congress’s intent, as expressed through its legislative decisions and statutory text, is clear: statistical sampling is off limits only when “determin[ing] [the] population for purposes of apportionment of Representatives.” 13 U.S.C. § 195. In every other context—including the redistricting context—statistical sampling is fair game. So the Census Act’s structure and drafting history disproves the thesis central to Plaintiffs’ legal theory: that the data underlying the tabulations contemplated in § 141(c) must be premised on the same methodology as those that underlie the tabulation contemplated in § 141(b). Rather, the Census Act itself demonstrates that the data underlying § 141(b) and § 141(c) may differ in methodology.

Plaintiffs also seem to argue in passing that the *Constitution* somehow obligates Defendants to produce redistricting data through their preferred methodology. Mot. 31. The single case Plaintiffs cite says nothing of the sort, and they quickly back away from

this undeveloped this argument. *See id.* (“At the very least, the constitutional question is raised”). But “[i]t is not enough merely to mention a possible argument in the most skeletal way, leaving the court to do counsel’s work, create the ossature for the argument, and put flesh on its bones.” *United States v. Zannino*, 895 F.2d 1, 17 (1st Cir. 1990); *see Forsberg v. Pefanis*, 634 F. App’x 676, 680 (11th Cir. 2015) (“Pefanis makes two other arguments, both of which he has forfeited by failing to develop them.”). In all events, Plaintiffs are mistaken. “[T]he constitutional purpose of the census” is “to determine the *apportionment* of the Representatives among the States,” *Wisconsin*, 517 U.S. at 20 (emphasis added)—that is, to determine the number of Representatives to which each State is entitled after the decennial census. Though “the States use the [census] results in drawing intrastate political districts,” that “consequence[]” is “not delineated in the Constitution.” *Id.* at 5–6 (emphasis added); *see also* Departments of Commerce, Justice, and State, The Judiciary, and Related Agencies Appropriations Act, 1998, § 209(a)(2), Pub. L. No. 105–119, 111 Stat. 2440 (1997) (codified at 13 U.S.C. § 141 note) (“1998 Appropriations Act”) (“[T]he *sole* constitutional purpose of the decennial enumeration of the population is the *apportionment* of Representatives in Congress among the several States.”) (emphases added).

Plaintiffs fare no better in attempting to import a judicially enforceable “accuracy” requirement into § 141(c).³ The decennial enumeration is an attempt to determine the true population of the United States, and “[t]hese figures may be as accurate as such immense undertakings can be.” *Gaffney*, 412 U.S. at 745. But as a matter of reality, census data “are inherently less than absolutely accurate.” *Id.* “Those who know about such

³ *Amica* Professor Bambauer argues that an accuracy requirement can be found in 13 U.S.C. § 181. Doc. 33 at 20–21. Even assuming that Professor Bambauer’s interpretation of § 181 were correct, § 181 expressly concerns certain data produced “[d]uring the *intervals between* each census of population required under section 141.” 13 U.S.C. § 181(a) (emphasis added). It does not relate to the data produced pursuant to § 141(c).

things,” the Supreme Court explained, “recognize this fact.” *Id.* And even if the enumeration could somehow result in a perfect population count, “the well-known restlessness of the American people means that population counts for particular localities are outdated long before they are completed.” *Karcher v. Daggett*, 462 U.S. 725, 732 (1983); *see also, e.g., Gaffney*, 412 U.S. at 745–46 (“[I]t makes little sense to conclude from relatively minor ‘census population’ variations among legislative districts that any person’s vote is being substantially diluted. The ‘population’ of a legislative district is just not that knowable to be used for such refined judgments.”).

In other words, the population counts determined in the decennial census are an approximation within a statistical range of the inherently unknowable population on Census Day. *See* Abowd Decl. ¶ 54. And the Census Bureau expects that the statistical “noise” that the differential-privacy algorithm will inject into those numbers will be measurably within that statistical range. *See id.* ¶¶ 54, 69. And in many cases, the post-differential-privacy population counts will have the effect of being *more* accurate.

For example, say the actual (but inherently unknowable) population of a given census block on Census Day is 50 individuals. The population count as determined by the actual enumeration might nonetheless record only 47 individuals as residing in the census block. But after the differential-privacy algorithm has been applied, the resulting population count increases by one person, *i.e.*, to 48 individuals. Plaintiffs’ legal position is that the post-differential-privacy population count of 48 individuals is illegally inaccurate while the 47-person figure is not—even though the 48-person figure is, in truth, more accurate. Such a result would not make sense.

Moreover, Plaintiffs’ position—that the Census Act incorporates *sub silentio* a judicially enforceable accuracy requirement hiding somewhere in the Census Act’s penumbrae, *see* Mot. 32–33—is the precise argument adopted by the district court in *National Urban League v. Ross* in enjoining the Secretary’s attempt to comply with the statutory apportionment deadline on the grounds that it was trumped by a supposed “statutory

duty of accuracy.” 489 F. Supp. 3d 939, 982, 994 (N.D. Cal. 2020), *stay denied in part*, 977 F.3d 770 (9th Cir. 2020), *stay granted*, 141 S. Ct. 18 (2020). We know how that ended: with a “rare and exceptional” Supreme Court stay. *Fargo Women’s Health Org. v. Schafer*, 507 U.S. 1013, 1014 (1993) (O’Connor, J., concurring in denial of stay application); *see Ross v. Nat’l Urban League*, 141 S. Ct. 18 (2020). And the Supreme Court granted the government’s requested stay despite the solo dissent’s position that “respondents [would] suffer substantial injury if the Bureau is permitted to sacrifice accuracy for expediency.” *Nat’l Urban League*, 141 S. Ct. at 21 (Sotomayor, J., dissenting).

“Through the Census Act, Congress has delegated its broad authority over the census to the Secretary.” *Wisconsin*, 517 U.S. at 19. And the Secretary and the Census Bureau—not Plaintiffs or the Court—are best positioned to optimally balance accuracy and confidentiality. Indeed, “there’s one branch Congress has not delegated any census decisions to: the judiciary.” *Nat’l Urban League*, 977 F.3d at 704 (Bumatay, J., dissenting). And just as “[t]here is no basis for the judiciary to inject itself into this sensitive political controversy and seize for itself the decision to reevaluate the competing concerns between accuracy and speed,” *see id.* at 713 (Bumatay, J., dissenting), there is similarly no basis for this Court to inject itself into the Census Bureau’s disclosure-avoidance methodology and seize for itself the decision to reevaluate the competing concerns between accuracy and confidentiality.

In sum, the Secretary will provide to the States redistricting data subject to differential privacy. Those data will be provided in a “tabulation,” and they represent the sub-state population. They are hence “tabulations of population.” 13 U.S.C. § 141(c). Because the Secretary will provide Alabama with “tabulations of population” as afforded to the State in § 141(c), “Defendants’ decision to apply differential privacy will” not “deprive Alabama of information which it is entitled to receive.” *Contra* Mot. 32. Alabama thus suffers no informational injury.

b. Sovereign Injury

Plaintiffs argue that the application of differential privacy will injure Alabama by “imped[ing] the State’s sovereign interest in drawing fair districts.” Mot. 33. In fact, Alabama will suffer no such injury for two independent reasons.

First, the redistricting data that the Secretary will ultimately produce to Alabama will be perfectly fit for redistricting. As explained above, the redistricting data need not exactly reflect the population counts from the enumeration, and the Census Bureau expects that the noise injected by differential privacy will be less than the estimated uncertainty inherent in the census. *See* Abowd Decl. ¶¶ 54, 69. After application of the differential-privacy algorithm, the redistricting data will remain “the best population data available” — indeed, Plaintiffs have not pointed to any other extant data that would be better — and, absent a source of better data, the redistricting data will constitute “the only basis for good-faith attempts to achieve population equality.” *Karcher*, 462 U.S. at 738.

Nonetheless, in an effort to show some sort of injury-in-fact, Plaintiffs contend — citing a short law journal article written by a law clerk — that if Alabama were to redistrict based on data subject to the differential-privacy algorithm, “litigation against the State” will be “especially likely.” Compl. ¶ 144. But Plaintiffs do not explain what source of alternative data could undergird such imagined lawsuits. And in all events, “[a]llegations of injury based on predictions regarding future legal proceedings are . . . too speculative to invoke the jurisdiction of an Article III Court.” *Platte River Whooping Crane Critical Habitat Maint. Tr. v. Fed. Energy Regulatory Comm’n*, 962 F.2d 27, 35 (D.C. Cir. 1992). Indeed, the Supreme Court has “been reluctant to endorse standing theories that require guesswork as to how independent decisionmakers will exercise their judgment.” *Clapper v. Amnesty Int’l USA*, 568 U.S. 398, 413 (2013). Moreover, injuries-in-fact must be “real, immediate, and direct.” *Ga. Republican Party v. SEC*, 888 F.3d 1198, 1202 (11th Cir. 2018). And “[a]lthough imminence is concededly a somewhat elastic concept, it cannot be

stretched beyond its purpose, which is to ensure that the alleged injury is not too speculative for Article III purposes—that the injury is *certainly* impending.” *Clapper*, 568 U.S. at 409 (emphasis in original). Alabama’s supposed injury—the possibility of future litigation brought by third parties on a speculative basis at some point in the distant future—cannot support standing.

Second, even if Alabama believes that it cannot use the redistricting data as produced by the Secretary, Alabama law does not obligate Alabama to use that data in drawing districts. “While the use of census data is the general practice, no stricture of the federal government requires States to use decennial census data in redistricting, so long as the redistricting complies with the Constitution and the Voting Right Act.” *Ohio v. Raimondo*, No. 3:21-cv-064, 2021 WL 1118049, at *8 (S.D. Ohio Mar. 24, 2021), *appeal filed*, No. 21-3294 (6th Cir. docketed Mar. 25, 2021); *see Burns v. Richardson*, 384 U.S. 73, 91 (1966) (“[T]he Equal Protection Clause does not require the States to use total population figures derived from the federal census as the standard by which this substantial population equivalency is to be measured.”); *Tucker v. U.S. Dep’t of Commerce*, 958 F.2d 1411, 1418 (7th Cir. 1992) (Posner, J.) (“[S]tates are not required to use census figures for the apportionment of their legislatures.”). Rather, States are required to use “the best population data available” to redistrict, *City of Detroit v. Franklin*, 4 F.3d 1367, 1374 (6th Cir. 1993)—and that data does not necessarily have to derive from the decennial census.

And, in fact, nothing in Alabama’s Constitution requires that the State use U.S. census data for its state legislative apportionment or redistricting. To be sure, Plaintiffs argue that the Alabama Constitution: (i) “requires that the State Legislature use the number of inhabitants, as reported by the Census Bureau, to apportion the seats in the State House and State Senate,” and (ii) obligates “[t]he Legislature [to] conduct legislative redistricting based on the Census Bureau’s tabulations.” Mot. 7 (citing Ala. Const. §§ 197–200). But neither proposition is correct.

First, Alabama’s Constitution expressly provides that the State’s apportionment need not necessarily be based on U.S. census data. Though section 198 provides that Alabama’s representatives shall be apportioned among the State’s counties “according to the number of inhabitants in them . . . as ascertained by the decennial census of the United States,” Ala. Const. § 198, section 201 – which Plaintiffs conspicuously neglect to mention – provides in part that if the decennial census is not “full and satisfactory” to the State, then “the legislature shall have the power at its first session after the time shall have elapsed for the taking of said census, to provide for an enumeration of all the inhabitants of this state, upon which it shall be the duty of the legislature to make the apportionment of representatives and senators.” Ala. Const. § 201. Plaintiffs allege that the Alabama Legislature’s “‘first session after taking the decennial census of the United States’ began February 2, 2021, and will adjourn May 30.” Compl. ¶ 71. And this very lawsuit reflects that in Alabama’s view, the decennial census is not “full and satisfactory” to the State. Accordingly, Alabama’s Legislature *is currently empowered* to conduct its own statewide census, after which “it shall be the duty of the legislature to make the apportionment of representatives and senators.” Ala. Const. § 201.

Second, no provision of Alabama’s constitution obligates “[t]he Legislature [to] conduct legislative redistricting based on the Census Bureau’s tabulations.” *Contra* Mot. 7 (citing Ala. Const. §§ 199–200). Sections 199 obligates the legislature to conduct a new apportionment of representatives “after each . . . decennial census.” Ala. Const. § 199. Section 200 obligates the legislature “to divide the state into as many senatorial districts as there are senators” “after each . . . decennial census.” Ala. Const. § 200. Neither section refers to – let alone requires – the *use* of U.S. census data. *See id.*

Simply put, nothing in Alabama’s constitution obligates the State to use census data to fulfill its “sovereign interest in drawing fair districts.” Mot. 33. Rather, if Alabama (incorrectly) believes that the future census redistricting data will be unsuitable for apportionment and redistricting, Alabama may conduct its own census. *See* Ala. Const.

§ 201. And in that case, Alabama’s decision *not* to conduct its own census is a classic “self-inflicted harm” that “does not amount to an ‘injury’ cognizable under Article III.” *Nat’l Family Planning & Reproductive Health Ass’n, Inc. v. Gonzales*, 468 F.3d 826, 831 (D.C. Cir. 2006).

The United States District Court for the Southern District of Ohio recently arrived at a similar conclusion. In *Ohio v. Raimondo*, the State of Ohio sued Defendants, arguing “that the Census Bureau’s plan to deliver redistricting data by September 30, 2021 is contrary to the deadlines established in 13 U.S.C. § 141(c).” *Ohio*, 2021 WL 1118049, at *6. Like Alabama here, Ohio argued that the September delivery date impeded its sovereign interests. But just like Alabama’s constitution, Ohio’s constitution also “contemplates ways in which redistricting can be accomplished in the absence of census data.” *Id.* Because Ohio’s laws were not actually “frustrated or rendered invalid by the delay in census data,” “[t]he absence of census data thus does not stop the state from implementing its constitutional scheme or otherwise impinge on its sovereign interests in effectuating its law.” *Id.* at *7. The same analysis applies here.

To be clear, Defendants are not suggesting that Alabama actually conduct its own census. To Defendants’ knowledge, Alabama has no such expertise. But Alabama’s constitution expressly empowers the State to conduct its own census if it is displeased with this year’s decennial census – and if Alabama’s census produces better data than the decennial census, Alabama may use its census to redistrict. Alabama’s concerted decision not to avail itself of its own constitutional powers is a classic self-inflicted injury that cannot support standing.

Pointing to *Karcher v. Daggett*, 462 U.S. 725 (1983), Plaintiffs also suggest – contrary to the Alabama constitution – that the decennial census “‘is the only basis for good-faith attempts to achieve population equality.’” Mot. 33 (quoting *Karcher*, 462 U.S. at 738). But Plaintiffs misread *Karcher*. “The Court in *Karcher* did not hold that the states must use census figures to reapportion congressional representation.” *City of Detroit*, 4 F.3d at

1374. “The Supreme Court merely reiterated a well-established rule of constitutional law: states are required to use the ‘best census data available’ or ‘the best population data available’ in their attempts to effect proportionate political representation.” *Id.* And “[i]f figures other than the census count are the best population data available, the Supreme Court did not, in *Karcher*, bar their use.” *Id.*

c. Federal Funding

Plaintiffs allege that “[d]ecennial census data are also used in many federal funding formulas that distribute federal funds to states and localities each year.” Compl. ¶ 148; *see generally id.* ¶¶ 148–158. But Plaintiffs conspicuously do not allege that Alabama is likely—let alone substantially likely—to suffer a *loss* of federal funds based on the application of differential privacy. Indeed, Plaintiffs make no effort to plausibly allege that the level of noise that the differential-privacy algorithm will inject into the future redistricting data will suffice to move the needle on even a *single* source of Alabama’s federal funding—let alone move the needle in a manner that will actually injure the State. Instead, Plaintiffs merely allege (in conclusory fashion) that purported funding variables “will be *affected* by differential privacy” and that such supposed “variance will directly *affect* the amount of federal funding Alabama and its citizens receive.” *Id.* ¶¶ 152, 158 (emphases added). Even assuming these naked allegations could surmount the plausibility threshold, they do not suffice to show substantial risk of *injury*.

In fact, Plaintiffs’ own expert strongly suggests that, to the extent that Alabama’s funding would be affected by differential privacy, it will result in a *windfall* to the State. Plaintiffs allege that “the rural population rate is a primary determinant of where federal spending is allocated.” Compl. ¶ 157. And Plaintiffs’ expert Dr. Barber opines that “[p]laces with fewer people (rural locations) . . . are more likely to be impacted” by the application of differential privacy—and the impact is (in his opinion) that rural areas would *gain* population: that “small [census] blocks, on average, get bigger” and “the largest blocks, on average, get smaller.” Barber Rep., Doc. 3–5, at 13–14; *see also id.* at 15

(quoting the State of Washington: ““There is a bias in the demonstration data that causes areas with small populations to get larger while areas with larger populations get smaller.””); *id.* (quoting the State of Utah: ““We observe that the population loss in our cities and towns are re-allocated to unincorporated, rural areas of the state.””).

In their motion, Plaintiffs also argue that differential privacy will result in the misallocation of federal funds. *See* Mot. 52–55. But like the challenge to the census rejected by the Supreme Court for lack of standing and ripeness in *Trump v. New York*, 141 S. Ct. 530 (2020), Plaintiffs’ supposed funding “injuries” are also “riddled with contingencies and speculation that impede judicial review.” *Id.* at 535. Plaintiffs’ “misallocation” arguments mirror the arguments improperly accepted by the *New York* district court. *See, e.g., New York v. Trump*, 485 F. Supp. 3d 422, 451 (S.D.N.Y. 2020) (“degraded census data jeopardizes various sovereign interests in allocating funds and administering public works through programs that rely on quality census data”), *vacated and remanded*, 141 S. Ct. 530 (2020). And though the Supreme Court’s dissenters argued that the *New York* plaintiffs’ predictions about the allocation of federal funds should be sufficient for standing purposes, *see* 141 S. Ct. at 540 (Breyer, J., dissenting), the majority rejected that argument. *See id.* at 536 (“The impact on funding is no more certain. According to the Government, federal funds are tied to data derived from the census, but not necessarily to the apportionment counts addressed by the memorandum. . . . Under that view, changes to the Secretary’s § 141(b) report or to the President’s § 2a(a) statement will not inexorably have the direct effect on downstream access to funds or other resources predicted by the dissent.”) (citation omitted).

Just as in *New York*, Plaintiffs’ allegations and arguments regarding a supposed “‘substantial risk’ of reduced . . . federal resources” “involve[] a significant degree of guesswork.” 141 S. Ct. at 535–36. But the future application of differential privacy, like the future application of the presidential memorandum at issue in *New York*, will not

“predictably change the count.” *Id.* at 536 (emphasis added). Accordingly, Plaintiffs’ “prediction about future injury [is] just that—a prediction.” *Id.*

d. Vote Dilution

Plaintiffs also argue that “[t]he Census Bureau’s decision to apply differential privacy . . . creates a substantial risk that” the individual plaintiffs “will have their votes in local, state, and federal elections diluted.” Mot. 36. But “injury results only to those persons domiciled in the under-represented voting districts.” *Wright v. Dougherty Cnty.*, 358 F.3d 1352, 1355 (11th Cir. 2004) (per curiam) (quoting *Fairley v. Patterson*, 493 F.2d 598, 603 (5th Cir. 1974)). Individuals who “have not suffered any harm or injury by the mal-apportioned voting districts” lack standing. *Id.*; see also, e.g., *Common Cause v. Rucho*, 279 F. Supp. 3d 587, 610 n.7 (M.D.N.C. 2018) (three-judge court) (“Plaintiffs in underpopulated districts lack standing to challenge a districting plan on one-person, one-vote grounds.”) (citing *Fairley*, 493 F.2d at 603–04), *vacated and remanded on other grounds*, 138 S. Ct. 2679 (2018).

The individual plaintiffs do not know how the future application of the differential-privacy algorithm will affect the population counts at any level of census geography. Indeed, each of them declares that they do not presently know, “and, in fact, may never know . . . if [their] vote is being weighed as equally as the vote of another voter in a neighboring district.” Williams Decl., Doc. 3–9, ¶ 12; see Green Decl., Doc. 3–10, ¶ 16 (substantially similar); Aderholt Decl., Doc. 3–11, ¶ 26 (substantially similar). At best, Plaintiffs’ argument reduces to the notion that the individual plaintiffs’ votes *may* be diluted. But the Supreme Court’s decisions “are consistent in recognizing a high standard for the risk-of-harm analysis.” *Muransky v. Godiva Chocolatier, Inc.*, 979 F.3d 917, 927 (11th Cir. 2020) (en banc). And “[a]llegations of possible future injury are not sufficient.” *Clapper*, 568 U.S. at 409. See, e.g., *Mont. Envtl. Info. Ctr. v. Stone-Manning*, 766 F.3d 1184, 1189 n.4 (9th Cir. 2014) (45% chance of harm “does not suffice to show a substantial risk”).

e. Section 209

Plaintiffs also assert injury based on the supposed violation of § 209 of the 1998 Appropriations Act, Pub. L. No. 105–119. *See* Mot. 36–38. No such injury exists.

Section 209 provides in part that “[a]ny person aggrieved by the use of any statistical method in violation of the Constitution or any provision of law . . . in connection with the 2000 or any later decennial census, to determine the population for purposes of the apportionment or redistricting of Members in Congress, may in a civil action obtain declaratory, injunctive, and any other appropriate relief against the use of such method.” 1998 Appropriations Act, § 209(b). Even assuming *arguendo* that Plaintiffs constitute such “person[s] aggrieved,” the Eleventh Circuit has made clear that “alleging a statutory violation is not enough to show injury in fact.” *Muransky*, 979 F.3d at 924. And *U.S. House of Representatives* demonstrates this principle in the § 209 context. In that case, the Supreme Court indicated that § 209 “eliminated . . . prudential concerns,” *see* 525 U.S. at 328 – and then proceeded to explain that “the only open justiciability question in this case is whether appellees satisfy the requirements of Article III standing.” *Id.* at 329. If a mere statutory violation of § 209 were sufficient to create Article III standing, the Court’s standing analysis, *see U.S. House of Representatives*, 525 U.S. at 329–34, would have been entirely unnecessary. *See also Muransky*, 979 F.3d at 928 (“A conclusory statement that a statutory violation caused an injury is not enough.”).

2. Plaintiffs Are Not Injured by Delayed Redistricting Data

Plaintiffs argue that Alabama is injured by the “delay in producing the population tables.” Mot. 55. “When the federal government prevents a State from applying state law,” they argue, “the State suffers an irreparable harm.” *Id.* (citing *Maryland v. King*, 133 S. Ct. 1, 3 (2012) (Roberts, C.J., in chambers)). But as explained above, Defendants are not preventing Alabama from complying with its own state law, because Alabama’s own constitution does not require census data for redistricting purposes.

Plaintiffs also argue that “delivering redistricting data on September 30 will also likely leave Alabama’s Boards of Registrars at most only four months for reassigning their respective counties’ registered voters to their correct precincts and districts,” yet “[t]he reassignments typically take up to six months.” Mot. 56. But the Boards of Registrars can get started right now with information that the Census Bureau has already provided to Alabama. *See, e.g.,* Whitehorne Decl. ¶¶ 10–12. And Plaintiffs’ declarant also makes clear that the State can “push[] back [its] primary election” by seven weeks. Helms Decl., Doc. 3–3, ¶¶ 14–15. In all events, this is just another way of saying that the 2020 decennial census is not “full and satisfactory” to the State of Alabama, thus empowering Alabama’s legislature to “provide for an enumeration of all the inhabitants” of the State. Ala. Const. § 201. In any event, Plaintiffs—citing the Helms declaration—argue that the Secretary’s September delivery of redistricting data “*will* result” in one or more harms. Mot. 56 (emphasis added). But the Helms declaration they cite is not so definitive. Rather, the Helms declaration states that “[r]equiring the Boards of Registrars and county commissions to complete the reassignment process on an abbreviated schedule *could* result in one or more” harms. Helms Decl., Doc. 3–3, ¶ 12 (emphasis added). This equivocal declaration cannot support standing: “threatened injury must be *certainly impending* to constitute injury in fact”; “[a]llegations of *possible* future injury are not sufficient.” *Clapper*, 568 U.S. at 409 (emphases in original).

Plaintiffs also argue that “the Bureau’s delay harms” Representative Aderholt “by effectively reducing by at least four months the amount of time [he] can spend campaigning and fundraising.” Mot. 56; *see also* Compl. ¶ 197. But “[t]o establish standing, an injury in fact must be concrete.” *Salcedo v. Hanna*, 936 F.3d 1162, 1167 (11th Cir. 2019) (footnote omitted). In turn, “[a] ‘concrete’ injury must be ‘*de facto*’; that is, it must actually exist.” *Id.* Representative Aderholt’s supposed injury does not meet this standard. Plaintiffs do not contend that these lost months will make it less likely for Representative Aderholt to win reelection. And it is clear why: delayed redistricting data affects *every*

candidate—not just Representative Aderholt. In fact, as the incumbent, Representative Aderholt is perhaps likely to *benefit* from a shorter campaign cycle. In all events, Representative Aderholt cannot be said to be injured by the delay in producing redistricting data.

B. Plaintiffs’ Alleged Injuries Are Not Traceable to Defendants’ Actions

1. Plaintiffs’ Alleged Injuries Cannot Be Traced to Defendants’ Plan to Use Differential Privacy

For similar reasons, Plaintiffs fail to establish the requisite “causal connection between” their alleged injuries and the actions they challenge—*i.e.*, they cannot show that any alleged injury is “fairly . . . trace[able]” to Defendants’ actions. *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992). Specifically, Plaintiffs have failed to show that their alleged injuries related to redistricting—*i.e.*, Alabama’s “sovereign interest in drawing fair districts” and the individual plaintiffs’ interest in not having their votes diluted—are traceable to Defendants. *See* Mot. 33, 36. The Supreme Court has explained in no uncertain terms that “[r]edistricting is primarily the duty and responsibility of the State,” *Abbott v. Perez*, 138 S. Ct. 2305, 2324 (2018), and “involves *choices* about the nature of representation with which [courts] have been shown no constitutionally founded reason to interfere,” *Burns*, 384 U.S. at 92 (emphasis added). “While the use of census data is the general practice, no stricture of the federal government requires States to use decennial census data in redistricting, so long as the redistricting complies with the Constitution and the Voting Rights Act.” *Ohio*, 2021 WL 1118049, at *8. Thus, in dismissing the State of Ohio’s recent lawsuit against Defendants, Judge Rose concluded that Ohio’s alleged injuries were not traceable to Defendants’ challenged actions, but rather Ohio’s “independent decision to create a state redistricting timeline without the flexibility to accommodate the COVID-19 pandemic.” *Id.*

Here, Alabama’s timetables do not even appear to be incompatible with a September 30, 2021, release of redistricting data. *See* Helms Decl., Doc. 3–3, ¶¶ 14–15 (conceding

that the State can “push[] back” its primary by seven weeks). And in all events, Plaintiffs’ claimed injuries here could only occur if the Alabama legislature declines to exercise its power, in the event that the U.S. decennial census is “not full and satisfactory,” “to provide for an enumeration of all the inhabitants of th[e] state.” Ala. Const. § 201. So any purported injury Alabama may suffer is “fairly . . . trace[able]” to the Alabama legislature’s independent decision to use U.S. census data and the State’s failure to adjust its own timetables, not “the challenged action of the defendant.” *Lujan*, 504 U.S. at 560.

Moreover, even if the Alabama legislature were required to use U.S. census data, Plaintiffs cannot demonstrate traceability because they cannot show that differential privacy will result in data that is less accurate when “compared to a feasible, alternative methodology,” *Nat’l Law Ctr. on Homelessness & Poverty v. Kantor*, 91 F.3d 178, 183 (D.C. Cir. 1996) (emphasis omitted), or that the difference between the two methodologies is sufficiently large to produce some kind of harm, *id.* at 185–86; *see also Franklin v. Massachusetts*, 505 U.S. 788, 802 (1992) (plurality) (challengers to the allocation of overseas employees among states had “neither alleged nor shown . . . that [they would] have had an additional Representative if the allocation had been done using some other source of ‘more accurate’ data” and accordingly did not have standing “to challenge the accuracy of the data used in making that allocation”). As noted above, Plaintiffs maintain that differential privacy will result in inaccurate numbers, but they have identified no other feasible, Census Act-compliant disclosure-avoidance methodology that would produce more accurate numbers. While Plaintiffs note that the Census Bureau has relied on other disclosure-avoidance methods in the past, Mot. 9–12, Dr. Abowd’s declaration explains in detail why those methods are not feasible for the 2020 Census. *See Abowd Decl.* ¶¶ 41–43, 50–51. Absent a feasible alternative, Plaintiffs cannot contend that any alleged inaccuracy is, in fact, “caused” by differential privacy.

2. Plaintiffs' Alleged Injuries Cannot Be Traced to Defendants' Delay in Producing Redistricting Data

Plaintiffs also fail to establish traceability for their purported injuries allegedly arising out of the Bureau's delay in producing redistricting data. Again, because redistricting is ultimately the responsibility of the State, Plaintiffs cannot show that their purported injuries are traceable to the challenged actions of Defendants, as opposed to the State's independent decisions. For this reason, the *Ohio* court recently dismissed Ohio's delay claim on traceability grounds, 2021 WL 1118049, at *8, and because the same analysis applies here, this Court should do the same.

Plaintiffs also cannot establish traceability because they identify no feasible alternative to producing redistricting data by September 30, 2021. Plaintiffs suggest in passing that the Bureau could have "attempted to deliver apportionment and redistricting numbers to different States 'on a flow basis,'" "prioritizing the States whose laws rely on timely receipt of census data." Mot. 47. But that would place Alabama last in line as its constitution affords the State an alternative path. *See* Ala. Const. § 201; *see generally* Part I.A.1.b. In all events, as the Whitehorne declaration explains, even if the Census Bureau prioritized Alabama's redistricting data to the detriment of the other 49 States, "it would not be able to deliver the data more than a few weeks earlier than a single national release"; "[t]he resulting data may have uncaught errors from [having] been rushed through review without the benefit of review of all States at once"; and it would "delay the release of data for the other 49 states." Whitehorne Decl. ¶¶ 29–30. Because there is no feasible alternative, Plaintiffs cannot contend that their alleged injuries are "caused" by any action by the Bureau.

C. Plaintiffs' Purported Injuries Are Not Redressable

An injury is redressable only if "a decision in a plaintiff's favor would 'significantly increase the likelihood' that [plaintiff] would obtain relief that directly redresses the injury that [plaintiff] claims to have suffered." *Lewis v. Governor of Ala.*, 944 F.3d 1287,

1301 (11th Cir. 2019) (en banc). Plaintiffs must demonstrate not only that they have suffered an injury that is traceable to Defendants, but also that “redress is likely ‘as a practical matter.’” *Jacobson v. Fla. Sec’y of State*, 974 F.3d 1236, 1255 (11th Cir. 2020) (quoting *Utah v. Evans*, 536 U.S. 452, 461 (2002)). Here, Plaintiffs cannot demonstrate that any of their alleged injuries would be redressed by an order enjoining Defendants from using differential privacy or requiring Defendants to produce redistricting data sooner than is possible.

1. Enjoining Differential Privacy Would Not Redress Plaintiffs’ Alleged Injuries

An order enjoining the Census Bureau from using differential privacy for the 2020 Census would not “significantly increase the likelihood” that Plaintiffs’ alleged injuries would be redressed. To the contrary, there is a significant likelihood that an order enjoining differential privacy would only make any alleged injuries worse. If the Court were to enjoin differential privacy, the Census Bureau would still need to comply with sections 8 and 9 of the Census Act, which prohibit Defendants from “disclos[ing] the information reported by, or on behalf of, any particular respondent,” or “mak[ing] any publication whereby the data furnished by any particular establishment or individual . . . can be identified.” 13 U.S.C. §§ 8(b), 9(a)(2). But the Census Bureau cannot rely solely on the disclosure avoidance methods used in the 2010 Census, which would also allow individual respondents’ data to be identified. *See* Abowd Decl. ¶¶ 38–39.

To comply with sections 8 and 9 of the Census Act, the Census Bureau would instead have to “swap” or “suppress” data at the census block level. *Id.* ¶¶ 40–43. This would *exacerbate* Plaintiffs’ alleged injuries, not redress them, because “[b]oth choices would delay results and diminish accuracy.” *Id.* ¶ 84. For example, Plaintiffs allege that differential privacy “impede[s] the State’s sovereign interest in drawing fair districts.” Mot. 33. As explained above, differential privacy will not cause any such injury to Alabama’s sovereign interests. *See supra*, Part I.A.1.b. By contrast, swapping or suppression

at the levels necessary to protect the census data could very well impede Alabama's ability to draw fair districts. *See* Abowd Decl. ¶¶ 42, 43, 87. Thus, "as a practical matter," an order enjoining differential privacy is not likely to redress Plaintiffs' claimed injuries resulting from allegedly inaccurate data. *Jacobson*, 974 F.3d at 1255.

An order enjoining the use of differential privacy would also only extend the Bureau's delay in providing redistricting data. As Dr. Abowd explains, it would take the Bureau "multiple months" to develop, test, and implement any alternative disclosure-avoidance methodology. Abowd Decl. ¶ 85. Accordingly, the relief that Plaintiffs seek—an order enjoining differential privacy—would hinder, rather than help, the Bureau's ability to produce redistricting data to the States as soon as possible.

2. Requiring the Census Bureau to Produce Redistricting Data Sooner Would Not Redress Plaintiffs' Alleged Injuries

Nor can Plaintiffs demonstrate redressability as to their delay claim. As Judge Rose observed in holding that the State of Ohio had not demonstrated redressability in its similar challenge to the Census Bureau's delay, "a judicial decree is only the means to an end: 'At the end of the rainbow lies not a judgment, but some action (or cessation of action) by the defendant that the judgment produces.'" *Ohio*, 2021 WL 1118049, at *5 (quoting *Doe v. DeWine*, 910 F.3d 842, 850 (6th Cir. 2018)). "In other words, '[r]edress is sought *through* the court, but *from* the defendant,' and '[t]he real value of the judicial pronouncement—what makes it a proper judicial resolution of a case or controversy rather than an advisory opinion—is in the settling of some dispute which affects the behavior of the defendant towards the plaintiff.'" *Id.* (quoting *Doe*, 910 F.3d at 850) (emphasis added).

Here, as in *Ohio*, "[Alabama] seeks an advisory opinion that cannot redress their claimed injury." *Id.*; *see also Jacobson*, 974 F.3d at 1255 (redress must be likely "as a practical matter"); *Brown v. Berhndt*, 12-cv-24-KGB, 2013 WL 1497784, at *5 (E.D. Ark. Apr. 10, 2013) (no standing where "injunctive relief [wa]s impossible"). That's because it is

“not possible under any scenario for the Census Bureau to produce these data at this time or at any time in the immediate future, and the Census Bureau would be unable to comply with any such order from the Court.” Whitehorne Decl. ¶ 14. “[T]he Census Bureau must complete a series of interim steps prior to delivering the redistricting data,” and “[e]ach of these interim steps, in order, is required to move to the next.” *Id.* ¶¶ 15–16. Those steps will likely not be completed until September 30, 2021, though the Bureau expects to be able to make a “legacy” format of the redistricting data file available to States in mid-to-late August. *Id.* ¶¶ 14–16, 27–28. Although the 2020 Census Operational Plan provided for only three months from the planned release of apportionment data on December 31, 2020, *see* Mot. 28, 49, the Bureau now requires five months because of operational changes that the Bureau made to expedite the release of the constitutionally required apportionment counts, including “decoupling” certain processes that the Bureau would have normally completed at the same time. Thieme Decl. ¶¶ 84–86.

Alabama’s purported injury is “also unredressable when it comes to redistricting for congressional (as opposed to state) elections.” *Ohio*, 2021 WL 1118049, at *5. In order to draw congressional districts, Alabama must first know the number of Representatives it will have in Congress to know how many districts to draw. 2 U.S.C. § 2c. But the Census Bureau has not yet finished, and neither the Secretary nor the President have yet reported, the apportionment of Representatives. Once the President reports the appointment numbers to Congress, apportionment will be entirely in Congress’s hands to accept or reject. *See* 2 U.S.C. § 2a(b) (commanding that apportionment only occurs “under [2 U.S.C. § 2a] or subsequent statute”). So even if the Court ordered the Census Bureau to produce redistricting data immediately, Alabama would be no closer to drawing congressional districts until Congress has determined the number of Representatives to which Alabama is entitled. In such circumstances, redressability (and standing) are lacking. *See Leifert v. Strach*, 404 F. Supp. 3d 973, 982 (M.D.N.C. 2019) (no redressability where

“[i]t is not merely speculative, but rather impossible, for the requested relief to remedy the alleged injury”).

Put simply, Alabama seeks the impossible. But “a court may not require an agency to render performance that is impossible.” *Am. Hosp. Ass’n v. Price*, 867 F.3d 160, 167 (D.C. Cir. 2017). Indeed, “[i]t has long been settled that a federal court has no authority . . . to declare principles or rules of law which cannot affect the matter in issue in the case before it.” *Church of Scientology of Cal. v. United States*, 506 U.S. 9, 12 (1992). The Court should therefore reject Alabama’s request for an advisory opinion based on the hypothetical world in which it were possible for the Census Bureau to comply with Alabama’s requested relief. The Court cannot “order a party to jump higher, run faster, or lift more than she is physically capable.” *Am. Hosp. Ass’n*, 867 F.3d at 168; Whitehorne Decl. ¶ 14 (explaining that “it would be a physical impossibility” to provide redistricting data at this time).

II. PLAINTIFFS ARE NOT ENTITLED TO A PRELIMINARY INJUNCTION.

“A preliminary injunction is an extraordinary remedy never awarded as of right.” *Winter v. Nat. Res. Def. Council, Inc.*, 555 U.S. 7, 24 (2008). Its “chief function . . . is to preserve the status quo until the merits of the controversy can be fully and fairly adjudicated.” *Ne. Fla. Chapter of Ass’n of Gen. Contractors of Am. v. City of Jacksonville*, 896 F.2d 1283, 1284 (11th Cir. 1990). But Plaintiffs are not asking the Court to preserve the status quo. Entering Plaintiffs’ proposed injunction would *upend* the status quo and would effectively constitute final relief in Plaintiffs’ favor by forcing the Census Bureau to completely overhaul its existing disclosure-avoidance methodology and to make wholesale, untested operational changes to produce redistricting data as quickly as possible.

Even assuming that Plaintiffs’ proposed relief could be characterized as a preliminary injunction, Plaintiffs do not satisfy any of the preliminary-injunction standards. “In order to obtain [a preliminary injunction], a party must establish four separate require-

ments — namely, that (1) it has a substantial likelihood of success on the merits; (2) irreparable injury will be suffered unless the injunction issues; (3) the threatened injury to the movant outweighs whatever damage the proposed injunction may cause the opposing party; and (4) if issued, the injunction would not be adverse to the public interest.” *Swain v. Junior*, 961 F.3d 1276, 1284–85 (11th Cir. 2020). And the latter two factors “merge when, as here, the Government is the opposing party.” *Id.* at 1293.

Plaintiffs “bear[] the burden of persuasion to clearly establish all . . . of these prerequisites.” *Wreal, LLC v. Amazon.com, Inc.*, 840 F.3d 1244, 1247 (11th Cir. 2016). “[F]ailure to meet even one dooms” Plaintiffs’ bid for a preliminary injunction. *Id.* at 1248.

A. Plaintiffs Are Unlikely to Succeed on the Merits of Their Differential Privacy Claims.

1. Plaintiffs’ Census Act Claim Is Not Likely to Succeed

Plaintiffs are not likely to prevail on their § 141(c) claim. *See* Compl. ¶¶ 198–202. As explained above, Defendants’ use of differential privacy will comply with § 141(c). *See supra* Part I.A.1.a.

Moreover, Alabama lacks a private right of action to assert a claim under § 141(c). “Like substantive federal law itself, private rights of action to enforce federal law must be created by Congress.” *Alexander v. Sandoval*, 532 U.S. 275, 286 (2001). “Where Congress has not created a private right of action, courts may not do so, ‘no matter how desirable that might be as a policy matter, or how compatible with the statute.’” *Bellitto v. Snipes*, 935 F.3d 1192, 1202 (11th Cir. 2019) (quoting *Sandoval*, 532 U.S. at 287).

The only private right of action to enforce § 141(c) flows through § 209(b) of the 1998 Appropriations Act.⁴ Section 209(b) provides a private right of action to “[a]ny person aggrieved by the use of any statistical method in violation of the Constitution or any

⁴ In their motion, Plaintiffs seem to suggest that § 209(b) provides them with a separate substantive claim. *See, e.g.*, Mot. 37–38 (“Defendants have violated Plaintiffs’

provision of law . . . in connection with the 2000 or any later decennial census, to determine the population for purposes of the apportionment or redistricting of Members in Congress.” Even assuming *arguendo* that differential privacy constitutes a “statistical method” as defined in § 209, Alabama is not a “person aggrieved.”

Section 209 states that “an aggrieved person . . . includes— (1) any resident of a State whose congressional representation or district could be changed as a result of the use of a statistical method challenged in the civil action; (2) any Representative or Senator in Congress; and (3) either House of Congress.” 1998 Appropriations Act § 209(d). Absent from this list of “aggrieved person[s]” are “States.” Plaintiffs nonetheless argue that the Court should infer that “Alabama is an ‘aggrieved person,’ too.” Mot. 37. But Congress did not include “States” in its list of “aggrieved persons,” and for this Court to do so would run counter to the “longstanding interpretive presumption that ‘person’ does not include the sovereign.” *Return Mail, Inc. v. U.S. Postal Serv.*, 139 S. Ct. 1853, 1861–62 (2019). For this reason, there is a “background presumption that States are not ‘persons.’” *Cook Cnty. v. United States ex rel. Chandler*, 538 U.S. 119, 133 n.10 (2003); *see Vt. Agency of Nat Res. v. United States ex rel. Stevens*, 529 U.S. 765, 780–88 (2000) (State is not a “person” for False Claims Act purposes). And “although the presumption is not a hard and fast rule of exclusion . . . it may be disregarded only upon some affirmative showing of statutory intent to the contrary.” *Return Mail, Inc.*, 139 S. Ct. at 1862.

If anything, the statutory text reflects Congress’s intent to *exclude* States from the definition of aggrieved persons. After all, this is not a situation where Congress left the term “person” undefined. Rather, Congress enacted a specific definition of “aggrieved

rights under Public Law No. 105–119, § 209(b).”). But Plaintiffs do not assert a claim for violation of § 209(b). *See generally* Compl. ¶¶ 198–241. And for good reason: Section 209(b) simply creates a private right of action. *See Common Cause v. Trump*, No. 1:20-cv-02023, -- F. Supp. 3d --, 2020 WL 8839889, at *12 (D.D.C. Nov. 25, 2020) (three-judge court); *Glavin v. Clinton*, 19 F. Supp. 2d 543, 547 (E.D. Va. 1998) (three-judge court), *aff’d sub nom.*, *Dep’t of Commerce v. U.S. House of Representatives*, 525 U.S. 316 (1999).

person” in § 209(d). That definition even included “either House of Congress” – hardly within the usual definition of “person.” But despite the Supreme Court’s “background presumption that States are not ‘persons,’” *Cook Cnty.*, 538 U.S. at 133 n.10, Congress – which is presumed to “legislate[] with knowledge of [the Supreme Court’s] basic rules of statutory construction,” *McNary v. Haitian Refugee Ctr., Inc.*, 498 U.S. 479, 496 (1991) – declined to include “States” in its definition of “aggrieved person.”

Plaintiffs acknowledge that States are “not expressly named in the statute,” but nonetheless have argued that “[t]he statute’s natural reading includes the States alongside Section 209(d)’s enumerated parties.” Pls. Mot., Doc. 2, at 5–7. Hardly. Given (i) the background presumption that “persons” do not include States, and (ii) Congress expressly included its Houses in defining “aggrieved person[s]” yet did not “expressly” include States, the “statute’s natural reading” is that “aggrieved person[s]” do not include “States.” Plaintiffs also argue that “a contrary interpretation would contravene the statute’s purpose.” Pls. Mot., Doc. 2, at 6. Even assuming Plaintiffs could be considered the arbiters of congressional purpose, “it is ultimately the provisions of our laws rather than the principal concerns of our legislators by which we are governed.” *Oncale v. Sundowner Offshore Servs., Inc.*, 523 U.S. 75, 79 (1998).

Nor can Plaintiffs rely on the fact that the “aggrieved person” is defined as “includ[ing]” various persons and entities. 1998 Appropriations Act § 209(d). After all, the Dictionary Act defines “person” as “includ[ing] corporations, companies, associations, firms, partnerships, societies, and joint stock companies, as well as individuals,” 1 U.S.C. § 1 (emphasis added) – yet the Supreme Court held that “[t]he absence of any comparable provision extending the term to sovereign governments implies that Congress did not desire the term to extend to them.” *United States v. United Mine Workers of Am.*, 330 U.S. 258, 275 (1947).

In sum, Alabama cannot take advantage of § 209's narrow right of action to enforce § 141(c), and in any event, none of the Plaintiffs are likely to succeed on their § 141(c) claims. *See supra* Part I.A.1.a.

2. The Individual Plaintiffs' Equal Protection Claim Is Not Likely to Succeed

The individual plaintiffs are not likely to succeed on their one-person-one-vote equal-protection claim. *See* Mot. 35–36. Only individuals residing in under-represented voting districts may bring one-person-one-vote claims. *Wright*, 358 F.3d at 1355. And “over-represented voting district members are barred from bringing suit on behalf of persons who reside in under-represented voting districts.” *Id.* Even assuming *arguendo* that census operational decisions could be susceptible to vote-dilution challenges, Plaintiffs have made clear that they do not know — “and, in fact, may never know” — whether their votes will be diluted. Williams Decl., Doc. 3–9, ¶ 12; Green Decl., Doc. 3–10, ¶ 16; Aderholt Decl., Doc. 3–11, ¶ 26. Plaintiffs concede that they cannot demonstrate any actual or impending vote dilution, and are thus unlikely to succeed on their vote-dilution claims.

Plaintiffs also have not pointed the Court to any case where census operations were enjoined on the grounds that resulting census data might lead States to redistrict in a manner that violated the one-person-one-vote principle. And, in fact, the Supreme Court has rejected such a bid. *See Wisconsin v. New York*, 517 U.S. 1, 16–17 (1996) (“[T]he ‘good-faith effort to achieve population equality’ required of a State conducting intrastate redistricting does not translate into a requirement that the Federal Government conduct a census that is as accurate as possible.”). This is not surprising. As explained above, “the Equal Protection Clause does not require the States to use total population figures derived from the federal census as the standard by which this substantial population equivalency is to be measured.” *Burns*, 384 U.S. at 91. Indeed, Alabama’s own constitution empowers the State to conduct its own census if it is dissatisfied with the decennial census. Ala. Const. § 201. So to the extent that the application of differential privacy

could be said to cause any “vote dilution,” the decision to use federal census data is Alabama’s alone, and no equal-protection claim may lie against the Defendants.

3. Plaintiffs’ APA Challenges to Differential Privacy Are Not Likely To Succeed

Plaintiffs’ APA claims face a fundamental problem: the Census Bureau has not yet finalized critical details on how it will use differential privacy. Plaintiffs acknowledge this. *See, e.g.*, Mot. 1 (describing differential privacy as a “still developing confidential algorithm”); Bryan Rep., Doc. 3–6, at 7 (claiming that “[t]he Census Bureau . . . will make a final decision about how DP will be implemented in the redistricting data by early May 2021”). The “in-progress” nature of differential privacy dooms Plaintiffs’ APA claim because this Court lacks jurisdiction when there is no final agency action. *See Nat’l Parks Conservation Ass’n v. Norton*, 324 F.3d 1229, 1236 (11th Cir. 2003).

Plaintiffs try to get around this problem by styling their legal theory as a facial challenge to differential privacy, basing their claim on the 2018 Operational Plan that announced the Census Bureau intended to use differential privacy but that left the critical details to be filled in later. *See* Mot. 40. But the core of Plaintiffs’ concerns relate to the Census Bureau’s later and still ongoing choices like setting the specific privacy-loss budget. And in any event, even if Plaintiffs’ claims (APA or otherwise) were proper and could be characterized as a facial challenge to the 2018 Operational Plan, they would run headlong into the doctrine of laches. *See infra* Part II.A.4.

a. The Differential Privacy Announcement Was Not Final Agency Action

No “agency action” as defined by the APA. A cognizable APA claim must challenge a “circumscribed, discrete agency action[]” and it cannot advance a “broad programmatic attack” on an agency’s operations. *Norton v. S. Utah Wilderness All.*, 542 U.S. 55, 61–62 (2004) (“SUWA”); *see also* 5 U.S.C. § 551; 5 U.S.C. § 701(b)(2) (agency action includes “an agency rule, order, license, sanction, relief, or the equivalent or denial thereof”). Put differently, the APA does not permit a plaintiff to attack an agency program “consisting

of . . . many individual actions” simply by characterizing it as “agency action” under the APA. *Lujan v. Nat’l Wildlife Fed’n*, 497 U.S. 871, 893 (1990). While “[c]ourts are well-suited to reviewing specific agency decisions,” they are “woefully ill-suited [] to adjudicate generalized grievances asking [them] to improve an agency’s performance or operations.” *City of New York v. U.S. Dep’t of Def.*, 913 F.3d 423, 431 (4th Cir. 2019).

The Census’s data-processing operations, including disclosure avoidance, “expressly are tied to one another,” so altering any of these operations “would impact the efficacy of the others, and inevitably would lead to court involvement in ‘hands-on’ management of the Census Bureau’s operations.” *NAACP v. Bureau of the Census*, 945 F.3d 183, 191 (4th Cir. 2019) (citing *SUWA*, 542 U.S. at 66–67), *aff’g in part and rev’g in part*, 399 F. Supp. 3d 406 (D. Md. 2019); *see, e.g.*, Whitehorne Decl. ¶¶ 15–16, 21; Abowd Decl. ¶¶ 84–89. In *NAACP*, plaintiffs challenged certain “design choices” within the Census Bureau’s December 2018 Operational Plan—the same Plan that Plaintiffs here claim was the “final agency action” by announcing that the Bureau intended to use differential privacy. *Compare NAACP*, 945 F.3d at 187–88 n.1 with Compl. ¶ 79 n.6. The *NAACP* district court found that the design choices within the Operational Plan were not agency action, explaining that “if the Court were to interject itself into the Bureau’s process during the critical final preparations, requiring—as Plaintiffs request—its monitoring and approval of the plans along the way, it is hard to imagine that this oversight would not hinder the process as opposed to facilitate it.” *NAACP v. Bureau of the Census*, 382 F. Supp. 3d 349, 372 (D. Md. 2019).

Plaintiffs’ differential privacy challenge fails this same threshold agency-action inquiry because it is a “broad programmatic attack” on the Census Bureau’s disclosure avoidance operations, not a challenge to “circumscribed, discrete agency action[.]” *SUWA*, 542 U.S. at 61–62. While Plaintiffs style their legal theory as a facial challenge to differential privacy, a close read of their complaint, motion, and expert reports shows

they ask the Court to scrutinize highly technical policy decisions related to how the Census Bureau *might* implement differential privacy. For example, Plaintiffs take issue with what data will remain untouched during the disclosure-avoidance operations — data sets known as “invariants.” Compl. ¶ 89; Mot. 14. They complain that the planned 2020 invariants include “(1) the total population of each State, (2) the total housing units at the census block level, and (3) the number of group quarters facilities by type at the census block level.” Mot. 14 & n.30 (citing a February 2021 summary file). But the 2020 invariants were not finalized in the 2018 Operational Plan and thus are beyond the scope of Plaintiffs’ current APA claims.

The Census Bureau’s policy choices for what data to hold constant when applying differential privacy could have dominoing impacts on both the disclosure avoidance process and the interrelated data-processing steps. *See* Abowd Decl. ¶ 88. So any Court order commanding the Bureau to set particular invariants — or an order changing to a different disclosure-avoidance method altogether — would require “a sweeping overhaul to the [processing operations], which exceeds the scope of reviewable ‘agency action.’” *NAACP*, 399 F. Supp. at 422. Plaintiffs’ requested relief shows the challenged action is not the type of circumscribed agency action that the APA makes reviewable.

No jurisdiction because no final agency action. Even if the 2018 decision to use differential privacy constitutes agency action, this Court still lacks jurisdiction over Plaintiffs’ APA claims because that decision was not *final* agency action. *See In re MDL-1824 Tri-State Water Rights Litig.*, 644 F.3d 1160, 1181, 1185 (11th Cir. 2011). To demonstrate subject-matter jurisdiction, Plaintiffs must show that “the administrative action in question is [] ‘final’ within the meaning of 5 U.S.C. § 704.” *Nat’l Parks Conservation Ass’n v. Norton*, 324 F.3d at 1236. To be final agency action, the challenged action must “mark the ‘consummation’ of the agency’s decision-making process — it must not be of a merely tentative or interlocutory nature” and the challenged action “must be one by which rights or obligations have been determined, or from which legal consequences will flow.” *Bennett v.*

Spear, 520 U.S. 154, 177–78 (1997); *Tri-State Water Rights*, 644 F.3d at 1181. Plaintiffs fail on both counts.

First, the Supreme Court has held that interim decisions about Census data processing are not complete until the final decision-maker delivers the data. In *Franklin*, Massachusetts challenged a particular method to assign home states for military personnel stationed abroad. *Franklin*, 505 U.S. at 790. The Supreme Court rejected Massachusetts’ challenge, explaining that there was no final agency action until the President delivered the final apportionment count to Congress pursuant to Section 141(b). 505 U.S. at 800. The interim steps taken by the Secretary of Commerce and the Census Bureau prior to the delivery of the final apportionment numbers under § 141(b) were tentative and not final agency action. *Id.*; *see id.* at 799 (“The President, not the Secretary, takes the final action that affects the States.”). The same analysis applies to the redistricting under § 141(c); the interim steps taken by the Census Bureau before the Secretary delivers the redistricting data to the states cannot constitute final action. *See City of Detroit*, 4 F.3d at 1377 n.6. Final action will occur only when the Secretary delivers the final data to the States, which has not yet occurred. Plaintiffs’ contrary position—that the Census Bureau’s operational plan can somehow bind the Secretary of Commerce—has no merit. “There is no authority for the proposition that a lower component of a government agency may bind the decision making of the highest level.” *Cnty. Care Found. v. Thompson*, 318 F.3d 219, 227 (D.C. Cir. 2003).

Even setting aside *Franklin*, the factual issues that Plaintiffs flag in their motion and declarations underscore why there is no final agency action. Plaintiffs and their declarants flag potential issues in non-final, *demonstration* data products—not the final redistricting data. *See generally* Mot. 20–24; Bryan Rep., Doc. 3–6. The entire point of releasing the demonstration products was to identify issues like the ones flagged by Plaintiffs. *See* Abowd Decl. ¶¶ 58–61. Census Bureau officials have explained that they are still working to resolve issues like those identified in the motion and declarations. *See*

id. ¶¶ 68–71. In these circumstances where the agency is actively working to resolve known issues, this court should follow the instruction of the Eleventh Circuit, “exercise restraint,” and let the Census Bureau use “its own institutional expertise” to address potential issues before releasing its final product. *LabMD, Inc. v. FTC*, 776 F.3d 1275, 1278 (11th Cir. 2015) (no final agency action when “agency proceeding is ongoing”).

Critical details of how the Census Bureau will implement differential privacy have not yet been finalized. In particular, the privacy-loss budget will not be set until June. Abowd Decl. ¶ 71. Plaintiffs acknowledge that the eventual privacy-loss budget will affect the ultimate redistricting data: “Dialing the [privacy-loss budget] up to infinity results in perfect accuracy but theoretically imperfect privacy, whereas setting the [privacy-loss budget] at zero results in perfect privacy but useless data.” Mot. 13. And Plaintiffs recognize that the Census Bureau has not reached a final decision on this critical matter. See Mot. 40 (“To be sure, the Bureau has yet to set the privacy loss budget it will use—that decision is still in the works.”) (emphasis added); *id.* at 1 (“the Bureau intends to provide numbers produced by a *still developing* confidential algorithm”) (emphasis added); *id.* at 17 (the Bureau “seeks to impose a *still-developing theory of privacy* onto the decennial census”) (emphasis added). Plaintiffs’ expert, Mr. Bryan, was even more blunt: “The Census Bureau . . . *will make a final decision* about how DP will be implemented in the redistricting data by early May 2021.” Bryan Rep., Doc. 3–6, at 7 (emphasis added). The 2018 Operational Plan was not the consummation of decision-making; in many ways, it was just the beginning of a iterative process that is still in progress.

Second, even if the 2018 Operational Plan could somehow be considered the consummation of an agency’s decision-making, it is still not “final” under the APA because it does not “determine any rights or obligations and imposes no legal consequences.” *Clayton Cnty. v. FAA*, 887 F.3d 1262, 1266–67 (11th Cir. 2018). The Operational Plan’s announcement that the Census Bureau would use differential privacy was “purely infor-

mational,” “[c]ompell[ed] no one to do anything,” and “had no binding effect whatsoever – not on the agency and not on” the general public. *See Indep. Equip. Dealers Ass’n v. EPA*, 372 F.3d 420, 427 (D.C. Cir. 2004).

The decision to use differential privacy, standing alone, does not cause the purported “legal consequences” claimed by Plaintiffs. Citing no case law, Plaintiffs claim that the 2018 decision to use differential privacy causes legal consequences by supposedly impeding Alabama’s ability to redistrict and creating a “substantial risk” that individual plaintiffs’ constitutional rights will be abridged. Mot. 40. But those purported “legal consequences” do not inherently flow from the use of differential privacy; those purported consequences flow from third-party decisions regarding redistricting – such as Alabama’s decision not to conduct the census for which its own constitution allows. And even if legal consequences flow from the final redistricting data, that final product will depend on the Census Bureau’s ultimate methodology and privacy-loss budget – not the 2018 decision to use differential privacy.

Nor do the supposed accuracy issues flagged by Plaintiffs somehow demonstrate that the decision to use differential privacy had legal consequences. Plaintiffs’ analysis was based on preliminary demonstration data. As Plaintiffs acknowledge, “the Bureau has stated that it intends to set a less conservative privacy loss budget for the final tabulations of population than it did for the demonstration products.” Mot. 35. And thus the final redistricting “numbers will be less skewed than they are in the demonstration data.” *Id.* Until the Census Bureau sets the final privacy-loss budget and releases the final numbers, Plaintiffs have not shown that there will be *any* legal consequences from differential privacy. The mere announcement that the Census Bureau would use differential privacy lacks legal consequence and is not reviewable final agency action under the APA.

b. Even Assuming the Differential Privacy Announcement Constituted Final Agency Action, It Did Not Violate the APA

Plaintiffs argue that “[t]he Census Bureau’s decision to adopt differential privacy is contrary to law, contrary to constitutional right, and in excess of statutory authority.” Mot. 40. They premise this argument on the notion that “the *application* of differential privacy to the population tabulations given to the States” is somehow inconsistent with 13 U.S.C. § 141(c) or that it would supposedly “create a substantial risk that individual Plaintiffs will have their equal protection rights violated.” Mot. 40 (emphasis added).

But Plaintiffs cannot challenge the eventual *application* of differential privacy through an APA challenge to the decision to ultimately implement *some form* of differential privacy. Indeed, Plaintiffs’ § 141(c) and equal-protection challenges are premised on the notion that the Census Bureau’s eventual application of differential privacy will not hold sub-state population counts invariant. But, as explained above, the invariants were not finalized in the 2018 Operational Plan and thus are beyond the scope of Plaintiffs’ current APA challenges to the 2018 Operational Plan. And even assuming *arguendo* that the 2018 Operational Plan had finalized invariants for the eventual application of differential privacy, Plaintiffs’ facial APA challenge to that supposed decision still would fail, as Plaintiffs are not likely to succeed on their § 141(c) or equal-protection claims. *See generally supra* Parts I.A.1.a, I.A.1.d, II.A.1, II.A.2.

For similar reasons, Plaintiffs cannot demonstrate that the *decision* to adopt differential privacy is arbitrary and capricious. Plaintiffs hinge their arbitrary-and-capricious APA claim on the notion that the application of differential privacy will supposedly preclude the Secretary from meeting her obligations “to report accurate tabulations of population under subsection 141(c),” Mot. 42—that is, Plaintiffs’ complaint is again about invariants, and not the disclosure-avoidance methodology in the abstract. And as the 2018 Operational Plan did not declare that sub-state population counts would be made variant, any such decision cannot be challenged in Plaintiffs’ APA claim.

And in all events where (unlike here) there is final agency action, the arbitrary and capricious standard is “exceedingly deferential.” *Sierra Club v. Van Antwerp*, 526 F.3d 1353, 1360 (11th Cir. 2008). The Court is “not authorized to substitute [its] judgment for the agency’s as long as its conclusions are rational.” *Miccosukee Tribe of Indians of Fla. v. United States*, 566 F.3d 1257, 1264 (11th Cir. 2009). “A court simply ensures that the agency has acted within a zone of reasonableness and, in particular, has reasonably considered the relevant issues and reasonably explained the decision.” *FCC v. Prometheus Radio Project*, 141 S. Ct. 1150, 1158 (2021). And the Eleventh Circuit “believe[s] it appropriate to give an extreme degree of deference to the agency when it is evaluating scientific data within its technical expertise.” *Nat’l Mining Ass’n v. Sec’y, U.S. Dep’t of Labor*, 812 F.3d 843, 866 (11th Cir. 2016).

As explained *supra*, Background Parts C & D, the Census Bureau determined that the disclosure-avoidance methodologies it previously used to protect census data were no longer sufficient given the rise in computing power, and that differential privacy was “[t]he best disclosure avoidance option that offers a solution capable of addressing the new risks of reconstruction-abetted re-identification attacks, while preserving the fitness-for-use of the resulting data for the important governmental and societal uses of census data.” Abowd Decl. ¶ 47. The Census Bureau’s decision-making process is not arbitrary or capricious.

Plaintiffs’ arbitrary-and-capricious claim is premised on a number of false notions. For starters, Plaintiffs argue that “the Bureau has not shown that traditional disclosure avoidance methods like data swapping are insufficient to meet” the Census Act’s confidentiality requirements. Mot. 41–42. But that position is easily rebutted by the JASON report that Plaintiffs repeatedly cited in their opening brief. *E.g.*, JASON, *Formal Privacy Methods for the 2020 Census* (Apr. 2020) at 6, available [here](#) (“Approaches to disclosure avoidance such as swapping and top and bottom coding applied at the level used in the

2010 census are insufficient to prevent re-identification given the ability to perform database reconstruction and the availability of external data.”); *accord, e.g.,* Abowd Decl. ¶¶ 38–39, 41–43, 50.

Plaintiffs further argue that the Census Bureau “misinterpreted the confidentiality requirements of § 9,” contending that “[c]onfidentiality is only implicated—in theory—when a recipient of census data uses the information published by the Bureau *together with* other datasets” to re-identify respondents. Mot. 43 (emphasis in original). But it is Plaintiffs that misconstrue the Census Act’s confidentiality requirements. Initially, Plaintiffs’ argument fails the plain text of the statute. Section 9(a) provides that Bureau staff, among others, generally may not “make any publication whereby the data furnished by any particular establishment or individual under this title can be identified.” 13 U.S.C. §§ 9(a), (a)(2). And the Census Bureau demonstrated, as corroborated by JASON, that the 2010 disclosure-avoidance methodology resulted—given recent advances in computing power—in publications that allowed respondent data to be identified. Indeed, under Plaintiffs’ atextual reading of § 9, the Census Bureau need not apply *any* disclosure-avoidance mechanism at all—not even to protect the sole, easily-identifiable Filipino American in the 20-person census block in the data-swapping example they provide, *see* Mot. 10–11—because, in their view, the Census Bureau would only violate § 9 if the *Bureau* publishes respondents’ names and addresses.

In all events, Plaintiffs conspicuously ignore § 9’s companion, 13 U.S.C. § 8, as well as on-point Supreme Court precedent. In *Baldrige v. Shapiro*, 455 U.S. 345 (1982), the Supreme Court expressly rejected the argument that the Census Act’s “confidentiality provisions protect raw data only if the individual respondent can be identified.” *Id.* at 355. Rather, “Congress plainly contemplated that raw data reported by or on behalf of individuals was to be held confidential and not available for disclosure.” *Id.*; *see also id.* at 361 (“§ 8(b) and § 9(a) of the Census Act embody explicit congressional intent to preclude *all*

disclosure of raw census data reported by or on behalf of individuals”) (emphasis in original). So while re-identification may not be possible without the use of other sources of data, the Census Bureau’s database-reconstruction experiment demonstrated that its 2010 census publications could be reverse-engineered, and thus resulted in an unfortunate “disclosure of raw census data reported by or on behalf of individuals.” *Id.* at 361.

Nor did Defendants ignore their end-users’ reliance interests. The 2018 Operational Plan itself made clear that the application of differential-privacy constitutes “a delicate balancing act”: “enough noise must be added to protect confidentiality, but too much noise could damage the statistic’s fitness-for-use.” 2018 Operational Plan, Doc. 3–4, at 140. “The Census Bureau decided that differential privacy was the best tool after analyzing the various options through the lens of economics.” Abowd Decl. ¶ 41. “Efficiently protecting privacy can be viewed as an economic problem because it involves the allocation of a scarce re-source – confidential information – between two competing uses: public data products and privacy protection.” *Id.* The Bureau’s “empirical analysis showed that differential privacy offered the most efficient trade-off between privacy and accuracy – our calculations showed that the efficiency of differential privacy dominated traditional methods.” *Id.* “In other words, regardless of the level of desired confidentiality, differential privacy will always produce more accurate data than the alternative traditional methods considered by the Census Bureau.” *Id.*

The ultimate accuracy of the redistricting data will also be much greater than the demonstration data released to date. By April 30, 2021, the Census Bureau will release a further set of demonstration data that employs a higher privacy-loss budget, tuned for accuracy, and which “better approximates the final privacy-loss budget that will likely be selected for the redistricting data product.” Abowd Decl. ¶ 69. Plaintiffs and their experts will have at least four weeks to review the next set of demonstration data, perform their analyses, and submit feedback before DSEP sets the final privacy-loss budget

and production parameters in June. See U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#).

Finally, even assuming that the 2018 Operational Plan could be said to violate the APA § 706(2), see Compl. ¶¶ 210–218, the only remedy would be to “set [it] aside” and “remand [it] to the agency for additional investigation.” 5 U.S.C. § 706(2); *Fla. Power & Light Co. v. Lorion*, 470 U.S. 729, 744 (1985). Indeed, under APA § 706(2), “it is not a court’s role to direct the agency how to act. Rather, a court’s role is to review the agency’s decision and, if it cannot be sustained, remand to the agency.” *Neto v. Thompson*, -- F. Supp. 3d --, 2020 WL 7310636, at *11 (D.N.J. Dec. 10, 2020) (citing *Dep’t of Homeland Sec. v. Regents of the Univ. of California*, 140 S. Ct. 1891, 1907–08 (2020)). And any such remand would add “multiple months” of further delay. Abowd Decl. ¶ 85; see generally *supra*, Background Part D.

4. The Doctrine of Laches Bars Plaintiffs’ Differential Privacy Claims

Assuming the Court concludes that Plaintiffs are bringing a facial challenge to the 2018 Operational Plan (as opposed to a challenge to the application of differential privacy, which would be premature), such a challenge is barred by the doctrine of laches. The doctrine of laches “protect[s] defendants against unreasonable, prejudicial delay in commencing suit.” *SCA Hygiene Prods. Aktiebolag v. First Quality Baby Prods., LLC*, 137 S. Ct. 954, 960 (2017). The doctrine provides defendants with an equitable defense that warrants consideration “separate from a statute of limitations [defense].” *Grayson v. Allen*, 499 F. Supp. 2d 1228, 1236 (M.D. Ala.), *aff’d*, 491 F.3d 1318 (11th Cir. 2007). The doctrine “will bar a claim when three elements are present: (1) a delay in asserting a right or a claim; (2) that the delay was not excusable; and (3) that there was undue prejudice to the party against whom the claim is asserted.” *Venus Lines Agency, Inc. v. CVG Int’l Am., Inc.*, 234 F.3d 1225, 1230 (11th Cir. 2000); see also *Wood v. Raffensperger*, No. 1:20-CV-04651–

SDG, -- F. Supp. 3d --, 2020 WL 6817513, at *7 (N.D. Ga. Nov. 20, 2020), *aff'd*, 981 F.3d 1307 (11th Cir. 2020). All three elements are easily satisfied here.

First, Plaintiffs have delayed considerably in asserting their claims. Plaintiffs acknowledge that the Bureau announced its decision to use differential privacy for the 2020 Census “in September 2017” and added differential privacy to the 2020 Census Operational Plan “in December 2018.” Mot. 39. Under this theory, Plaintiffs knew or should have known the facts giving rise to their claims by December 2018 at the latest. Rather than timely bringing their claims once Plaintiffs became aware of the Bureau’s plans, however, Plaintiffs waited years to bring their lawsuit, until after the Bureau had already begun processing data and is now on the verge of releasing data in a matter of months. This years-long wait undoubtedly counts as a “delay.” *See, e.g., Wood*, 2020 WL 6817513, at *7 (laches barred challenge to November 2020 election where plaintiff was aware of basis for claim as early as March 2020); *Stone v. U.S. Postal Serv.*, 383 F. App’x 873, 875 (11th Cir. 2010) (laches barred claim due to plaintiffs’ three-year delay).

Second, Plaintiffs’ delay is inexcusable. Plaintiffs take the position that the Census Bureau’s December 2018 operational plan constitutes final agency action that is “ripe for review.” Mot. 39–40. Given that position, there is no excuse for waiting more than two years to challenge that decision. To be sure, the Bureau continues to refine its differential-privacy algorithm, and has not yet set the privacy-loss budget. But in Plaintiffs’ view, that decision is “immaterial” to their claims because “by definition, *any* application of differential privacy will produce erroneous numbers.” *Id.* at 35, 40 (emphasis added). Plaintiffs identify no reason in either their complaint or their motion why they waited until the eleventh hour to file suit. Indeed, Alabama *did* file suit against the Census Bureau in 2018 over the Bureau’s “Residence Rule” — a suit that remains pending in the Northern District of Alabama. *See Compl., Alabama v. Dep’t of Commerce*, No. 18-cv-772 (N.D. Ala. May 21, 2018). But Alabama waited until March 2021 to bring any challenge

to the Bureau's plan to use differential privacy, despite their claim that "any" application of differential privacy would be unlawful.

Third, Plaintiffs' delay has unduly prejudiced Defendants. If Plaintiffs had brought their challenge when the Census Bureau announced it would be using differential privacy, the Bureau would have had ample time to implement any operational consequences of an adverse decision before releasing redistricting data to the states. Now, with post-processing operations well underway and the release of data fast approaching, an adverse decision would significantly disrupt the Bureau's completion of the census. As Dr. Abowd explains, it would take "multiple months" to develop, test, and implement an alternative disclosure methodology. Abowd Decl. ¶ 85. Changing course at the last minute also poses significant risks to the accuracy of the data. See Thieme Decl. ¶ 74. Moreover, by bringing suit now during what is the busiest time of the decade for the Census Bureau, Plaintiffs have subjected the Bureau to the significant and unnecessary burden of having to defend against a federal lawsuit seeking to upend its entire framework for ensuring privacy while simultaneously working to complete the actual census itself. All of this could have been avoided if Plaintiffs had not delayed in bringing their claims.

B. Plaintiffs' Challenge to the February 12 Press Release Is Not Likely to Succeed.

Plaintiffs bring two statutory challenges to the Bureau's February 12 Press Release announcing that it would release redistricting data by September 30, 2021: (i) a claim that the press release "violates the Census Act," Mot. 44–45; Compl. ¶¶ 219–22, and (ii) a claim that the press release violates the APA, Mot. 46–50; Compl. ¶¶ 223–27. Neither challenge is likely to succeed.

1. Plaintiffs' Claim that the Press Release "Violates the Census Act" Is Not Likely to Succeed

Plaintiffs are unlikely to succeed on their claim that the February 12 Press Release violates § 141(c) of the Census Act. As an initial matter, Plaintiffs lack a private right of

action to bring this claim. As noted, the only private right of action to enforce § 141(c) flows through § 209(b) of the 1998 Appropriations Act. But that section provides a private right of action only to certain statutorily defined “aggrieved persons” to challenge “the use of any statistical method in violation of the Constitution or any provision of law . . . to determine the population for purposes of the apportionment or redistricting of Members in Congress.” And none of the Plaintiffs can use § 209 to challenge the February 12 Press Release because § 209 allows for challenges only to “statistical methods,” and the press release is obviously not a “statistical method.”⁵ Plaintiffs argue that the February 12, 2021 Press Release was “likely” a “byproduct of its . . . decision to implement differential privacy,” which Plaintiffs contend is a “statistical method.” See Pls. Mem., Doc. 2, at 4–5. But Plaintiffs are wrong as a factual matter — as the Thieme declaration explains, the “creation of the [Microdata Detail File] is not the reason that the Census Bureau will be unable to meet the statutory deadline.” See Thieme Decl. ¶ 71. Indeed, the Bureau has allotted approximately three weeks to apply differential privacy, while the disclosure-avoidance procedures used in the 2010 census took nearly four weeks. *Id.* And, more fundamentally, § 209 does not allow for challenges to press releases that are alleged “by-product[s]” of a statistical method — whatever that means. It allows only for challenges to statistical methods themselves.

Plaintiffs thus have no cause of action under the Census Act or § 209 to pursue an alleged violation of the statutory deadline in § 141(c). Nor is there any other basis for Plaintiffs to pursue this claim. While federal courts may “in some circumstances” grant injunctive relief against officials who are alleged to have violated federal law, “[t]he power of federal courts of equity to enjoin unlawful executive action is subject to express and implied statutory limitations.” *Armstrong v. Exceptional Child Ctr., Inc.*, 575 U.S. 320,

⁵ Additionally, as explained above, Alabama is not an “aggrieved person” under the statute, and so Alabama could not take advantage of § 209(b)’s narrow cause of action to enforce § 141(c) in any event. See *supra* Part II.A.1.

326–27 (2015). By expressly authorizing a cause of action for “aggrieved persons” to bring claims challenging “statistical methods” —but *only* statistical methods—Congress impliedly limited plaintiffs’ ability to challenge other alleged violations of the Census Act. *See id.* at 328 (holding that Medicaid Act foreclosed equitable relief because “sole remedy” Congress provided for in statute was for Secretary to withhold funds).

Nor is review available under the “*ultra vires*” doctrine or any other purported nonstatutory basis for review. Review under the *ultra vires* doctrine “is essentially a Hail Mary pass — and in court as in football, the attempt rarely succeeds.” *Nyunt v. Broad. Bd. of Governors*, 589 F.3d 445, 449 (D.C. Cir. 2009) (Kavanaugh, J.). Among other requirements, a plaintiff must show that the agency’s error is “so extreme that one may view it as jurisdictional or nearly so.” *Id.* (quoting *Griffith v. Fed. Labor Relations Auth.*, 842 F.2d 487, 493 (D.C. Cir. 1988)); *see also Protect Our Parks, Inc. v. Chicago Park Dist.*, 971 F.3d 722, 728 (7th Cir. 2020) (plaintiffs must show that defendants acted “beyond their legal authority”). Plaintiffs have not even attempted to make that showing here. Plaintiffs do not argue that the Census Bureau lacks the statutory authority to report tabulations of population after the deadline has passed, so *ultra vires* review does not even apply. And even if it did, Plaintiffs cannot show that the agency’s error was “so extreme” as to be “jurisdictional or nearly so,” where the Bureau could not meet the statutory deadline due to extraordinary events outside its control.

Finally, even if Alabama had a cause of action under the statute or otherwise, injunctive relief would be inappropriate because, as noted, it is physically impossible for the Bureau to produce redistricting data at this time or any time in the immediate future. A court may not exercise its equitable powers to “require an agency to render performance that is impossible.” *Am. Hosp. Ass’n*, 867 F.3d at 167.

2. **Alabama’s APA Challenge to the February 12 Press Release Is Not Likely to Succeed**

Alabama is likewise unlikely to succeed under the APA because its claim does not challenge any final agency action. Alabama’s claim focuses exclusively on the Bureau’s February 12 Press Release and related blog post. Mot. 44–45 (citing Mot. Exs. 7 & 8). But, as explained above, final agency action occurs when the Secretary reports the final redistricting numbers. *See* Part II.A.3.a.; *Franklin*, 505 U.S. at 790; *City of Detroit*, 4 F.3d at 1377 n.6. So the Press Release is not final agency action reviewable under the APA.

a. **The February 12 Press Release Was Not Final Agency Action**

As explained above, final agency action “must mark the consummation of the agency’s decision-making process—it must not be of a merely tentative or interlocutory nature” and “must be one by which rights or obligations have been determined, or from which legal consequences will flow.” *Bennett*, 520 U.S. at 177–78. A cognizable APA claim must also challenge a “circumscribed, discrete agency action[]”; it cannot advance a “broad programmatic attack” on an agency’s operations. *SUWA*, 542 U.S. at 61–62. Alabama’s challenge to the February 12 Press Release satisfies none of the requirements for final agency action.

No Consummation of the Decisionmaking Process. To determine whether an agency action is final, “[t]he core question is whether the agency has completed its decisionmaking process.” *Franklin*, 505 U.S. at 797. The APA does not allow a party to challenge “preliminary, procedural, or intermediate agency action” until the agency completes its action. *See Nat’l Parks Conservation Ass’n*, 324 F.3d at 1236 (quoting 5 U.S.C. § 704).

As explained above, the Supreme Court has held that there is no final agency action until the President delivers the final apportionment count to Congress. *See Franklin*, 505 U.S. at 797. The interim steps taken by the Secretary of Commerce and the Census Bureau prior to the delivery of the final apportionment numbers are tentative and not final agency action. *Id.* Although *Franklin* dealt with apportionment, the same analysis

applies to the redistricting context. *See City of Detroit*, 4 F.3d at 1377 n.6 (relying on *Franklin*'s reasoning to conclude that "the Secretary's reporting of the [redistricting] counts for these purposes is a final agency action"). Since reporting of final redistricting data is reviewable final agency action, the tentative actions and decisions leading up to the delivery of the redistricting data are not reviewable under the APA.

Even setting aside this Supreme Court precedent, a press release explaining that the Census expects to deliver redistricting data by a certain date did not consummate anything; it simply provided a snapshot in time of the expected delivery date that had shifted over the past year due to many factors, including disruptions from COVID, wildfires, hurricanes, court orders, and issues in data processing. *See supra* Background Part E. The February 12 Press Release simply updated Census's estimated timeline, and of course, estimates can still change as data processing continues. *See Whitehorne Decl.* ¶ 17. The Press Release thus does not reflect any definitive decision at all.

No Legal Consequences. The February 12 Press Release is also not final agency action because it did not change any legal rights or have any legal consequences. *See Cal. Cmty's. Against Toxics v. EPA*, 934 F.3d 627, 638 (D.C. Cir. 2019) (no final agency action where "no direct and appreciable legal consequences" and no party "can rely on it as independently authoritative in any proceeding"). The February 12 Press Release did not change any rights or obligations: the Secretary will deliver redistricting data to the States, including Alabama, when the data becomes available. Like the 2018 Operational Plan, the Press Release was also "purely informational"; "[c]ompelling no one to do anything," the Press Release "had no binding effect whatsoever—not on the agency and not on" the general public. *Indep. Equip. Dealers Ass'n*, 372 F.3d at 427. And, as discussed above, Alabama faces no legal consequences if it does not receive redistricting data by the statutory deadline. *See generally supra* Part I.A.1.b. In fact, Alabama faces no legal consequences at all,

regardless of timing, because its own law fully contemplates how to accomplish apportionment and redistricting in the absence of what it considers to be “full and satisfactory” census data. *See* Ala. Const. § 201; *Ohio*, 2021 WL 1118049, at *6.

Improper Programmatic Attack. Finally, Alabama’s challenge to the February 12 Press Release fails the final-agency-action inquiry because it is a “broad programmatic attack” on the Census Bureau’s operations, not a “circumscribed, discrete agency action[.]” *SUWA*, 542 U.S. at 61–62. While “[c]ourts are well-suited to reviewing specific agency decisions,” they are “woefully ill-suited [] to adjudicate generalized grievances asking [them] to improve an agency’s performance or operations.” *City of New York*, 913 F.3d at 431. But that is exactly what Alabama seeks here. Because the Census Bureau’s data-processing operations are all interdependent and interrelated, *see, e.g.*, Thieme Decl. ¶ 5; Whitehorne Decl. ¶¶ 15–16, 21, producing redistricting data on a different timeline would require “a sweeping overhaul to the [processing operations], which exceeds the scope of reviewable ‘agency action.’” *NAACP*, 399 F. Supp. 3d at 422. Indeed, like the Census Bureau’s field operations, its data-processing operations “expressly are tied to one another,” so altering any of these operations “would impact the efficacy of the others, and inevitably would lead to court involvement in ‘hands-on’ management of the Census Bureau’s operations.” *NAACP*, 945 F.3d at 191 (citing *SUWA*, 542 U.S. at 66–67). That is “precisely the result that the ‘discreteness’ requirement of the APA is designed to avoid.” *Id.* (citing *SUWA*, 542 U.S. at 67).

b. The February 12 Press Release is Not Arbitrary or Capricious

Nor can Alabama demonstrate that the February 12 Press Release is arbitrary or capricious in violation of the APA. Where (unlike here) there is final agency action, the arbitrary and capricious standard is “exceedingly deferential.” *Sierra Club*, 526 F.3d at 1360. The Court is “not authorized to substitute [its] judgment for the agency’s as long as its conclusions are rational.” *Miccosukee Tribe of Indians of Fla.*, 566 F.3d at 1264. And

this Court should “give an extreme degree of deference to the agency when it is evaluating scientific data within its technical expertise.” *Nat’l Mining Ass’n*, 812 F.3d at 866; *see also Ranchers Cattlemen Action Legal Fund v. Dep’t of Agric.*, 415 F.3d 1078, 1093 (9th Cir. 2005) (“Deference to the informed discretion of the responsible federal agencies is especially appropriate, where, as here, the agency’s decision involves a high level of technical expertise.”).

Here, there is a reasoned explanation for the Secretary’s inability to transmit redistricting data by the statutory deadline: “[I]t is not possible under any scenario for the Census Bureau to produce these data at this time or any time in the immediate future.” Whitehorne Decl. ¶ 14. Nor can the Bureau’s delivery of redistricting data for all States at once be considered arbitrary or capricious. *Contra* Mot. 47. Even if the Census Bureau prioritized Plaintiff’s redistricting data to the detriment of the other 49 States, “it would not be able to deliver the data more than a few weeks earlier than a single national release”; “[t]he resulting data may have uncaught errors from [having] been rushed through review without the benefit of review of all States at once”; and it would “delay the release of data for the other 49 states.” Whitehorne Decl. ¶¶ 29–30.

Finally, even assuming that the February 12 Press Release could be considered “arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law,” the only remedy would be to “set [it] aside” and “remand [it] to the agency for additional investigation.” 5 U.S.C. § 706(2); *Fla. Power & Light Co.*, 470 U.S. at 744. Indeed, under the APA § 706(2), “it is not a court’s role to direct the agency how to act. Rather, a court’s role is to review the agency’s decision and, if it cannot be sustained, remand to the agency.” *Neto*, 2020 WL 7310636, at *11 (citing *Regents of the Univ. of California*, 140 S. Ct. at 1907–08). And while the Census Bureau would take any such remand seriously, it

would not change the fact that “it is not possible under any scenario for the Census Bureau to produce these data at this time or any time in the immediate future.” Whitehorne Decl. ¶ 14.⁶

C. Plaintiffs Will Suffer No Harm, Much Less Irreparable Harm.

“A showing of irreparable injury is the sine qua non of injunctive relief.” *Siegel v. LePore*, 234 F.3d 1163, 1176 (11th Cir. 2000) (en banc) (per curiam). And “the asserted irreparable injury must be neither remote nor speculative, but actual and imminent.” *Id.* “Issuing a preliminary injunction based only on a possibility of irreparable harm is inconsistent with [the Supreme Court’s] characterization of injunctive relief as an extraordinary remedy that may only be awarded upon a clear showing that the plaintiff is entitled to such relief.” *Winter*, 555 U.S. at 22. Here, Plaintiffs cannot establish that they will likely suffer irreparable harm as a result of either the Bureau’s use of differential privacy or its February 12 Press Release.

1. Plaintiffs Have Not Established Irreparable Harm Due to Differential Privacy

As a threshold matter, and assuming that the Court concludes that Plaintiffs are bringing a facial challenge to the 2018 Operational Plan (because any challenge to the application of differential privacy is premature), Plaintiffs’ unexplained delay in bringing their differential privacy claim undercuts their claim of irreparable injury. “[T]he very idea of a preliminary injunction is premised on the need for speedy and urgent action to protect a plaintiff’s rights before a case can be resolved on its merits.” *Wreal, LLC v. Amazon.com, Inc.*, 840 F.3d 1244, 1248 (11th Cir. 2016) (emphasis in original). “For this reason” federal courts “have found that a party’s failure to act with speed or urgency in

⁶ Contrary to Alabama’s protestations, Mot. 47, the Census Bureau *did* consider States’ self-imposed reliance on census-based redistricting data. As the Whitehorne declaration explains, however, “[w]ith the delay in the delivery of the redistricting data, there are now too many states (at least 27) to prioritize, in a fair, logical, and data-driven manner.” Whitehorne Decl. ¶ 26.

moving for a preliminary injunction necessarily undermines a finding of irreparable harm.” *Id.* Thus in *Wreal*, the Eleventh Circuit stated that “[a] delay in seeking a preliminary injunction of even only a few months—though not necessarily fatal—militates against a finding of irreparable harm.” *Id.*

The record here reflects Plaintiffs’ unexplained delay of at least two years. Plaintiffs represent in their motion that the Census Bureau announced its decision to use differential privacy in September 2017, and that the Census Bureau added differential privacy to its “fourth (and latest) version of the Bureau’s 2020 Census Operational Plan,” which was released in December 2018. Mot. 12. They reference demonstration data that the Census Bureau released in October 2019 and in May, September, and November of 2020 that, in their view, “have shown that differential privacy . . . inhibits a State’s right to draw fair lines.” *Id.* at 18. And though the Census Bureau continues to refine its differential-privacy algorithm and its various inputs, Plaintiffs’ position is that “by definition, *any* application of differential privacy will produce erroneous numbers.” *Id.* at 35 (emphasis added).

But Plaintiffs do not explain why they failed to bring a challenge shortly after the Census Bureau added differential privacy to its December 2018 operational plan. Nor do they explain why they didn’t bring such a challenge after the Census Bureau started releasing demonstration data in October 2019. Instead, for reasons they do not explain, Plaintiffs waited until March 2021 to file this suit and move for a preliminary injunction. “[A] party cannot delay . . . and then use an ‘emergency’ created by its own decisions concerning timing to support a motion for a preliminary injunction.” *Mortensen v. Mortg. Elec. Registration Sys., Inc.*, No. CV 09-0787-WS-N, 2010 WL 11425328, at *8 (S.D. Ala. Dec. 23, 2010). “[B]ecause the instant motion for preliminary injunction was filed not just months, but years, after the factual basis of the Plaintiffs’ claims were known to them, the Plaintiffs have not shown they will suffer imminent, irreparable harm.” *Thompson v. Merrill*, No. 2:16-cv-783-ECM, 2020 WL 3513497, at *3 (M.D. Ala. June 29, 2020) (Marks, C.J.).

Setting aside Plaintiffs' unexplained delay in bringing their claim, Plaintiffs also cannot demonstrate an irreparable injury because they have not demonstrated any injury at all. *See supra* Part I.A. Plaintiffs contend that they will suffer an irreparable injury because differential privacy will supposedly "make lawful redistricting difficult." Mot. 50. But, as explained above, the redistricting data that the Secretary produces will be perfectly suitable for redistricting. *See* Abowd Decl. ¶¶ 54–56, 65–66, 69. As Dr. Abowd explains, the latest demonstration data product that will be released by April 30 is "extremely accurate." *Id.* ¶ 54. For example, "[t]otal populations for counties have an average error of +/- 5 persons" (an error rate of about 0.04% of the counties' population), whereas "the average county-level estimation uncertainty of the census is +/- 960 persons (averaging 1.6% of the county census counts)." *Id.* "In the April 2021 Demonstration Data Product, Congressional districts as drawn in 2010 have a mean absolute percentage error of 0.06%." *Id.* ¶ 56. And the average state legislative district has an average error of 0.16% or less. *See id.* Such miniscule error cannot possibly interfere with Alabama's ability to "lawful[ly] redistrict[]" or "subject the State to the risk of litigation and liability." Mot. 50. And even if Alabama believed that it did, Alabama's constitution does not require it to use census data in drawing its districts. *See supra* Part I.A.1.b.

Nor have Plaintiffs demonstrated that differential privacy will impose irreparable "financial harm" on Alabama. *See* Mot. 52–55. Again, as explained above, Plaintiffs do not allege that Alabama is likely to suffer a *loss* of federal funds as a result of differential privacy, and make no effort to show that the level of noise that the differential-privacy algorithm will inject will affect any aspect of Alabama's federal funding. *See supra* Part I.A.1.c. To the contrary, Plaintiffs' own expert suggests that to the extent Alabama's funding would be affected by differential privacy at all, it would result in a windfall to the State because, he predicts, rural areas would tend to gain population. *Id.*

Moreover, even if Plaintiffs could establish some potential future injury, they cannot show that they are likely to suffer the kind of "imminent" irreparable harm that

would justify the extraordinary remedy of a preliminary injunction. *Wreal*, 840 F.3d at 1248. As explained above, the Census Bureau is still in the process of finalizing the differential privacy algorithm, and has not, for example, set the privacy-loss budget. *See supra* Background Part D. Until it does so, Plaintiffs cannot demonstrate that the amount of noise that differential privacy adds could possibly be so great as to cause the kinds of irreparable harms that Plaintiffs allege. *See* Mot. 50.

2. Plaintiffs Have Not Established Irreparable Harm on Their Delay Claim

Nor have Plaintiffs demonstrated that they will suffer irreparable harm if the Census Bureau releases redistricting data by September 30, 2021. *See* Mot. 55–56. Again, Plaintiffs have not demonstrated any harm at all, let alone irreparable harm. Plaintiffs’ claim to harm rests entirely on an assertion that Alabama will be unable to comply with its constitution but, as explained above, Alabama’s constitution does not require using decennial census data for redistricting where, as here, the State does not believe that data to be “full and satisfactory.” *See supra* Part I.A.1.b; Ala. Const. § 201. This case is therefore unlike *Maryland v. King*, 567 U.S. 1301 (2012) (Roberts, C.J., in chambers), where a portion of state law was enjoined, precluding the state from enforcing its provisions. *Id.* at 1303 (noting that inability to “employ a duly enacted statute” constitutes irreparable harm). Here, by contrast, Alabama’s constitution expressly contemplates a situation where census data would not be “full and satisfactory” to the State and affords its legislature an opportunity to conduct its own census. *See* Ala. Const. § 201. The realization of a circumstance expressly accounted for in a state’s law is not a frustration of that text or its purpose. *See Conn. Nat’l Bank*, 503 U.S. at 253–54 (courts “must presume that [the] legislature says in a statute what it means and means in a statute what it says there.”).

Alabama may well prefer to use census data for redistricting, but a frustration of an alleged preference, without a factual showing of likely real-world effects, is insufficient to constitute an irreparable injury. *Cf. Judicial Watch, Inc. v. U.S. Dep’t of Homeland*

Sec., 514 F. Supp. 2d 7, 10 (D.D.C. 2007) (“Although plaintiff’s desire to have its case decided in an expedited fashion is understandable, that desire, without more, is insufficient to constitute the irreparable harm[.]”). Were it otherwise, anyone that came to court with a preference for different census operations could obtain an injunction as a matter of course. That is not—and cannot be—the standard. *Siegel*, 234 F.3d at 1179 (“[P]roof of irreparable injury is an indispensable prerequisite to a preliminary injunction.”). And even assuming that Alabama would sustain likely real-world effects, the State has not explained why, unlike other States, *see supra* Background Part E, it cannot find a workable solution other than through this lawsuit.

Likewise, Plaintiffs cannot establish imminent irreparable harm based on the argument that delivering redistricting data by September 30 would leave Alabama’s Boards of Registrars with “only” four months to reassign voters to their correct precincts and districts. Mot. 56. Plaintiffs assert that four months will “likely” not be enough, *id.*, but the declaration that Plaintiffs cite does not support that assertion. *See* Helms Decl., Doc. 3–3, ¶¶ 5–15. The declaration states merely that in those counties that assign voters manually, the process “can” take “up to [six] months.” *Id.* ¶ 7. This statement appears to be based on one prior reassignment process in 2017 when local officials allegedly struggled to assign voters in six months. *Id.* ¶ 8. From this fact, the declarant infers that requiring officials to complete the reassignment process in four months instead of six “could” lead to increased costs, the “potential[.]” for mistaken reassignments, and the “potential[.]” for confusion. *Id.* ¶ 12. But such “remote [and] speculative” potential harms are insufficient to establish the “actual and imminent” harm necessary to justify a preliminary injunction. *Siegel*, 234 F.3d at 1176. Moreover, Plaintiffs’ declarant acknowledges that Alabama could simply move its 2022 primary election seven weeks to July 12, 2022, Helms Decl., Doc. 3–3, ¶¶ 14–15, which would give Alabama the six months that it says it needs to complete the reassignment process.

Finally, Representative Aderholt cannot establish irreparable harm based on the fact that the Bureau's delay "effectively reduc[es] by at least four months the amount of time [he] can spend campaigning and fundraising." Mot. 56. As explained above, delayed redistricting affects all candidates, and, as the incumbent, Representative Aderholt is perhaps even more likely to benefit from a shorter campaign cycle. *See supra* Part I.A.2. Thus, Representative Aderholt cannot demonstrate any injury at all, let alone an injury that is "actual and imminent."

D. Defendants and the Public Would Be Harmed by an Injunction.

Differential Privacy Is In The Public Interest. The harm to the government and the public would be severe if the Census Bureau were forced to abandon differential privacy. *See Swain*, 961 F.3d at 1293 (harm to opposing party and the public interest "merge" when relief is sought against the government).

Forcing the Census Bureau to develop a different disclosure-avoidance method would have cascading affects, including significant delay in releasing the redistricting data and decreased quality of the data ultimately released. The Census Bureau is in the final stages of planning how it will deploy differential privacy, which will be the culmination of a process that has been ongoing since at least 2017. Forcing the Bureau to change methods at this late hour would upend the schedule and cause significant delays—indeed, changing methods "would add significant additional time (at least several months) to the schedule for delivering redistricting data." Thieme Decl. ¶ 74. Since the Bureau announced that it would use differential privacy in 2017, States and other data users have provided "extensive actionable feedback" that "has informed ongoing [disclosure-avoidance] system improvements and design changes." U.S. Census Bureau, *2020 Disclosure Avoidance System Updates* (Feb. 23, 2021), available [here](#). Only one State—Alabama—has filed a lawsuit over the use of differential privacy. The other States deserve to get the data they expect without additional, undue delay caused by a preliminary injunction.

There is a strong public interest in protecting the confidentiality of census responses. The Supreme Court has recognized that “an accurate census depends in large part on public cooperation” and “[t]o stimulate that cooperation Congress has provided assurances that information furnished to the Secretary by individuals is to be treated as confidential.” *Baldrige*, 455 U.S. at 354. And a federal statute provides that that Census Bureau staff that publish information protected by 13 U.S.C. § 9 “shall be” subject to fines “or imprisoned not more than 5 years, or both.” 13 U.S.C. § 214.

The Census Bureau chose to use differential privacy because it is the best way to protect confidentiality while still providing quality, accurate redistricting data to the public. Other available disclosure-avoidance methods, including suppression or swapping, do not provide similarly powerful confidentiality protections, and “to achieve the necessary level of privacy protection, both enhanced data swapping and suppression [would have] severely deleterious effects on data quality and availability.” Abowd Decl. ¶ 51. And if the Bureau were nonetheless forced to provide detailed data at small geographic levels, it would expose the confidential information of millions of Americans who trusted the Bureau to keep their data secure.

The Census Bureau Cannot Provide Redistricting Data By March 31, 2021. It is now April, so it would be impossible for the Bureau to comply with any order requiring it to release redistricting data by March 31, 2021. Even an order requiring the Census Bureau to speed up the release of redistricting data faster than what Census Bureau officials have already announced would be difficult, if not impossible, to implement. Whitehorne Decl. ¶¶ 14–17, 21; *see supra* Part I.C. The Census Bureau’s current schedule reflects the realistic amount of time the Bureau has concluded it needs to complete the complex steps required to finish processing the various sources of data it received; verifying the quality of its tabulations; and preparing usable, accurate outputs that comply with statutory requirements for respondent confidentiality protection. Whitehorne Decl. ¶¶ 20–21, 28–30; Thieme Decl. ¶¶ 60–83 (detailing the steps that still need to be accomplished to deliver

redistricting data). An order requiring the Census Bureau to deliver data faster would yet again disrupt census operations, reduce the time for data quality checks, and make it even *more* difficult for the Census Bureau to complete its work. Whitehorne Decl. ¶¶ 28–30; Thieme Decl. ¶¶ 69, 73–74.

The harm from such a disruption would reverberate to other States and the public at large. If the Census Bureau were required to prioritize Alabama’s data, it may well have to delay delivery of other States’ data until past September 30, 2021. Whitehorne Decl. ¶¶ 30, 31. Such a delay would disrupt those other States’ redistricting plans—presumably leading those States to suffer the same kinds of harms Alabama alleges in this lawsuit. Already, at least one other state has brought a lawsuit like Alabama’s, requesting that its data be prioritized over those of other states. *See Ohio v. Raimondo*, No. 3:21-cv-064, 2021 WL 1118049 (S.D. Ohio Mar. 24, 2021), *appeal filed*, No. 21-3294 (6th Cir. docketed Mar. 25, 2021). Meanwhile, plaintiffs in California continue to assert that any shortening of data-processing operations would be unlawful. *See Nat’l Urban League v. Raimondo*, No. 20-cv-05799, ECF Nos. 465 & 467 (N.D. Cal. Feb. 3, 2021). The more courts intrude on census operations, the more entities will want to seek judicial intervention on their behalf, and the longer it will ultimately take to receive the results.

III. MANDAMUS RELIEF IS UNAVAILABLE.

In three short paragraphs, Plaintiffs argue that Alabama is entitled to “partial relief through a writ of mandamus requiring the Secretary to meet the statutory deadline of March 31 to deliver the tabulations of populations for redistricting to the States.” Mot. 58–59. “Mandamus is an extraordinary remedy which should be utilized only in the clearest and most compelling of cases.” *Cash v. Barnhart*, 327 F.3d 1252, 1257 (11th Cir. 2003). This is not that case. Plaintiffs’ bid to invoke the Mandamus Act, 28 U.S.C. § 1361, should be rejected.

“Under 28 U.S.C. § 1361, otherwise known as The Mandamus Act, the district court has original jurisdiction over a mandamus action to compel an officer or employee

of the United States or any agency thereof to perform a duty owed to the plaintiff.” *Cash*, 327 F.3d at 1257. “Mandamus relief is appropriate only when: (1) there is no other adequate remedy and (2) the plaintiff has a clear right to the relief requested (in other words, the defendant must have a clear duty to act).” *United States v. Salmona*, 810 F.3d 806, 811 (11th Cir. 2016). “Put another way, a writ of mandamus is intended to provide a remedy for a plaintiff only if he has exhausted all other avenues of relief and only if the defendant owes him a clear nondiscretionary duty.” *Id.* And “[a]lthough the issuance of a writ of mandamus is a legal remedy, it is largely controlled by equitable principles and its issuance is a matter of judicial discretion.” *Cash*, 327 F.3d at 1257–58; *see also, e.g., Lovitky v. Trump*, 949 F.3d 753, 759 (D.C. Cir. 2020) (“Even when the legal requirements for mandamus jurisdiction have been satisfied, however, a court may grant relief only when it finds compelling equitable grounds.”); Mot. 58 (acknowledging that “issuance of the writ” must be “‘appropriate under the circumstances’”) (quoting *Cheney v. United States Dist. Ct.*, 542 U.S. 367, 381 (2004)). Alabama is not entitled to mandamus relief for two independent reasons.

For starters, Alabama has not demonstrated a clear, mandatory duty that would afford it with a clear right to relief because “it is anything but clear that Congress intended the deadline[] at issue to be mandatory rather than directory.” *Friends of Aquifer, Inc. v. Mineta*, 150 F. Supp. 2d 1297, 1300 (N.D. Fla. 2001). Again, mandamus relief presupposes, *inter alia*, that “the defendant owes [the plaintiff] a clear nondiscretionary duty.” *Salmona*, 810 F.3d at 811. And “[f]or there to be a ‘duty owed to the plaintiff’ within the meaning of section 1361, there must be a mandatory or ministerial obligation. If the alleged duty is discretionary or directory, the duty is not ‘owed.’” *Maczko v. Joyce*, 814 F.2d 308, 310 (6th Cir. 1987). To be sure, as Plaintiffs point out, *see* Mot. 44–45, “the word ‘shall’ usually connotes a requirement.” *Maine Cmty. Health Options v. United States*, 140 S. Ct. 1308, 1320 (2020) (emphasis added). But, as the Supreme Court expressly noted, that is not always the case, and it is not the case here.

The *Friends of Aquifer* case is directly on point. That case concerned the Pipeline Safety Act, which provided in part that the Secretary of Transportation “shall prescribe standards” relating to certain hazardous liquid pipeline facilities by various dates certain. 150 F. Supp. 2d at 1298–99 (quoting Pipeline Safety Act, 49 U.S.C. § 60109). The Secretary allegedly did not discharge his statutory duties in that regard, and the plaintiff sought mandamus relief. *See id.* at 1298. Citing several cases, the court explained that “in a variety of contexts, courts have concluded that Congress’s use of the word ‘shall’ in directing the discharge of a specified duty does not require that the statute be construed as mandatory rather than directory.” *Id.* at 1300. The court noted that, like § 141(c) here, the Pipeline Safety Act neither imposed any “penalty or sanction for the Secretary’s failure to prescribe the requisite standards by the specified dates,” nor did it include any provision affording jurisdiction to plaintiffs “to compel the Secretary to prescribe certain standards required under the Act.” *Id.* at 1299–1300. Finding no “clear mandate from Congress that it intended the statutory deadlines at issue to be something other than directory, and absent a showing that Congress intended a clear right in Plaintiff to the relief sought,” the court declined to “exercise its equitable powers to order the Secretary to issue standards that are dependent upon technological complexities and developments that are peculiarly within the agency’s—not th[at] court’s—expertise.” *Id.* at 1301.

The same analysis applies here. Plaintiffs have not demonstrated any “clear mandate from Congress,” *id.*, that it intended the § 141(c) deadline to be mandatory rather than directory. To the contrary, there are no statutory consequences for missing the deadline, and historical practice supports the conclusion that census deadlines are directory in nature. And, like the *Friends of Aquifer* court, this Court should decline to “exercise its equitable powers” to order Defendants to rush the processing of the data Alabama seeks, which work is similarly “dependent upon technological complexities and developments that are peculiarly within” the Census Bureau’s expertise. *See Friends of Aquifer*, 150 F. Supp. 2d at 1301; *see also, e.g., Robertson v. Attorney General of U.S.*, 957 F. Supp. 1035, 1037

(N.D. Ill. 1997) (finding statutory deadline to be directory and declining to issue mandamus relief; “In order to achieve the goals of the statute, the Attorney General and INS may have to engage in lengthy investigations to determine the validity of a given marriage.”).⁷

Moreover, Alabama is not entitled to mandamus relief because, as explained above, the relief it seeks is impossible to provide. “[T]he writ of mandamus will not issue to compel the performance of that which cannot be legally accomplished.” *Am. Hosp. Ass’n*, 867 F.3d at 167. “[P]ossibility is a necessary and antecedent condition for the writ’s issuance.” *Id.* at 169 (collecting sources); see 52 Am. Jur. 2d § 24 (“Mandamus will not issue if the performance of the requested action is impossible”); 55 C.J.S. Mandamus § 19 (“The writ of mandamus will not lie where performance of the duty is impossible.”). Simply put, this Court “may not require” the Census Bureau “to render performance that is impossible.” *Am. Hosp. Ass’n*, 867 F.3d at 167.

This action plainly does not constitute the “the clearest and most compelling of cases” in which to invoke relief under the Mandamus Act. *Cash*, 327 F.3d at 1257. So Plaintiffs’ request for a writ of mandamus must be denied.

CONCLUSION

For the reasons explained above, Plaintiffs’ motion and petition should be denied.

⁷ Historical practice demonstrates that Congress considers census deadlines as directory. From the very first census, deadlines were missed for various reasons, but Congress either retroactively revised the statute to accommodate the late submission, or simply ignored that a deadline was missed. See An Act granting further Time for making Return of the Enumeration of the Inhabitants in the District of South Carolina, 1 Stat. 226 (1791). Congress likewise extended census deadlines throughout the 1800s whenever they were missed. See An Act to Extend the Time for Completing the Third Census, 2 Stat. 658 (1811); An Act to Amend the Act Entitled “An Act to Provide for Taking the Fourth Census,” 3 Stat. 643 (1821), An Act to Amend the Act for Taking the Fifth Census, 4 Stat. 439 (1831), An Act to Amend the Act Entitled “An Act to Provide for Taking the Sixth Census,” 5 Stat. 452 (1841), An Act Supplementary to the Act Entitled “An Act Providing for the Taking of the Seventh and Subsequent Censuses,” 9 Stat. 445 (1850).

DATED: April 13, 2021

Respectfully submitted,

BRIAN M. BOYNTON
Acting Assistant Attorney General

ALEXANDER K. HAAS
Director, Federal Programs Branch

BRAD P. ROSENBERG
Assistant Director, Federal Programs Branch

/s/ Elliott M. Davis
ZACHARY A. AVALLONE
ELLIOTT M. DAVIS (N.Y. Reg. No. 4596755)
JOHN ROBINSON
Trial Attorneys
Civil Division, Federal Programs Branch
U.S. Department of Justice
1100 L St. NW
Washington, DC 20005
Phone: (202) 514-5336
E-mail: elliott.m.davis@usdoj.gov

Counsel for Defendants

CERTIFICATE OF SERVICE

I hereby certify that on April 13, 2021, I filed with the Court and served on opposing counsel through the CM/ECF system the foregoing document.

DATED: April 13, 2021

/s/ Elliott M. Davis
ELLIOTT M. DAVIS
(N.Y. Reg. No. 4596755)
Trial Attorney
Civil Division, Federal Programs Branch
U.S. Department of Justice
1100 L St. NW
Phone: (202) 514-4336
Fax: (202) 616-8470
E-mail: elliott.m.davis@usdoj.gov

Counsel for Defendants

**IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION**

STATE OF ALABAMA, *et al.*,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF
COMMERCE, *et al.*,

Defendants.

Case No. 3:21-CV-211-RAH-ECM-KCN

DECLARATION OF JOHN M. ABOWD

I, John M. Abowd, make the following Declaration pursuant to 28 U.S.C. § 1746, and declare that under penalty of perjury the following is true and correct to the best of my knowledge:

BACKGROUND

1. I am the Chief Scientist and Associate Director for Research and Methodology at the United States Census Bureau. I have served in this capacity since June 2016. My statements in this declaration are based on my personal knowledge or on information supplied to me in the course of my professional responsibilities.
2. I received my Ph.D. in economics from the University of Chicago with specializations in econometrics and labor economics in 1977 (M.A. 1976). My B.A. in economics is from the University of Notre Dame.
3. I have been a university professor since 1976 when I was appointed assistant professor of economics at Princeton University. I was also assistant and associate professor of econometrics and industrial relations at the University of Chicago Graduate School of Business. In 1987, I was appointed associate professor of industrial and labor relations with indefinite tenure at Cornell University where I am currently the Edmund Ezra Day Professor. I am on unpaid leave from Cornell University to work in my current position at the Census Bureau as part of the Career Senior Executive Service.
4. I am a member and fellow of the American Association for the Advancement of Science, American Statistical Association, Econometric Society, and Society of Labor Economists (president 2014). I am an elected member of the International Statistical Institute. I am also a member of the American Economic Association, International Association for Official Statistics, National Association for Business Economists, American Association for Public Opinion Research, Association for Computing Machinery, and American Association of Wine Economists. I regularly attend and present papers at the meetings of these organizations.

5. I have served on the American Economic Association Committee on Economic Statistics. I have also served on the National Academy of Sciences Committee on National Statistics, the Conference on Research in Income and Wealth Executive Committee, and the Bureau of Labor Statistics Technical Advisory Board for the National Longitudinal Surveys (chair: 1999-2001).
6. I have worked with the Census Bureau since 1998, when the Census Bureau and Cornell University entered into the first of a sequence of Intergovernmental Personnel Act agreements and other contracts. Under those agreements, I served continuously as Distinguished Senior Research Fellow at the Census Bureau until I assumed my current position as Chief Scientist in 2016, under a new Intergovernmental Personnel Act contract. Since March 29, 2020, I have been in the Associate Director position at the Census Bureau as a Career Senior Executive Service employee.
7. From 2011 until I assumed my position as Chief Scientist at the Census Bureau in 2016, I was the lead Principal Investigator of the Cornell University node of the NSF-Census Research Network, one of eight such nodes that worked collaboratively with the Census Bureau and other federal statistical agencies to identify important theoretical and applied research projects of direct programmatic importance to the agencies. The Cornell node produced the fundamental science explaining the distinct roles of statistical policymakers and computer scientists in the design and implementation of differential privacy systems at statistical agencies.
8. I have published more than 100 scholarly books, monographs, and articles in the disciplines of economics, econometrics, statistics, computer science, and information science. I have been the principal investigator or co-principal investigator on 35 sponsored research projects. I was a founding editor of the [Journal of Privacy and Confidentiality](#) – an interdisciplinary journal, and I continue to serve as an editor and on the governance board. My full professional resume is attached to this report as Appendix A.

9. I have worked on and managed Census Bureau projects that were precursors to the Census Bureau's current program to implement differential privacy for the 2020 Census of Population and Housing. I was one of three senior researchers who founded the Longitudinal Employer-Household Dynamics (LEHD) program at the Census Bureau, which is generally acknowledged as the Census Bureau's first 21st Century data product: built to the specifications of local labor market specialists without additional survey burden, and published beginning in 2001 using state-of-the-art confidentiality protection via noise infusion. This program produces detailed public-use statistical data on the characteristics of workers and employers in local labor markets using large-scale linked administrative, census, and survey data from many different sources. In 2008, my work with LEHD led to the first production implementation worldwide of differential privacy as part of a product of the LEHD program called OnTheMap. The LEHD program also implemented other prototype systems to protect confidential information, including allowing the public to access synthetic micro-data confirmed via direct analysis of the confidential data on validation servers. A differentially private version of this system is under development at the Census Bureau but not for use with the 2020 Census.

IMPORTANCE OF CONFIDENTIALITY

10. Though participation in the census is mandatory under 13 U.S. Code § 221, in practice, the Census Bureau must rely on the voluntary participation of each household in order to conduct a complete enumeration.
11. One of the most significant barriers to conducting a complete and accurate enumeration are individuals' concerns about the confidentiality of census data. The Census Bureau's pre-2020 Census research showed that 28% of respondents were "extremely concerned" or "very concerned" and a further 25% were "somewhat concerned"

about the confidentiality of their census responses.¹ These concerns are even more pronounced in minority populations and represent a major operational challenge to enumerating traditionally hard-to-count populations.²

12. To secure voluntary participation, Congress first established confidentiality protections for individual census responses in the Census Act of 1879. These confidentiality protections were later expanded and codified in 13 U.S. Code §§ 8(b) & 9, which prohibits the Census Bureau from releasing “any publication whereby the data furnished by any particular establishment or individual under this title can be identified[,]” and allows the Secretary to provide aggregate statistics so long as those data “do not disclose the information reported by, or on behalf of, any particular respondent[.]” Title III of the Foundations for Evidence Based Policymaking Act of 2018 also requires statistical agencies to “protect the trust of information providers by ensuring the confidentiality and exclusive statistical use of their responses.”³
13. The broader scientific community generally concurs about the importance of rigorous protection of confidentiality by statistical agencies. For example, the National Academy of Sciences’ definitive guidebook for federal statistical agencies states “Because virtually every person, household, business, state or local government, and organization is the subject of some federal statistics, public trust is essential for the continued effectiveness of federal statistical agencies. Individuals and entities providing data di-

¹ U.S. Census Bureau (2019) “2020 Census Barriers, Attitudes, and Motivators Study Survey Report” <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2020-report-cbams-study-survey.pdf>, p.38-39.

² Ibid, p.39-42.

³ Title III of the Foundations for Evidence Based Policymaking Act of 2018, § 3563.

rectly or indirectly to federal statistical agencies must trust that the agencies will appropriately handle and protect their information.”⁴ The report also notes that respondents expect statistical agencies not to “release or publish their information in identifiable form.”⁵ The National Academies also broadly exhort statistical agencies to “continually seek to improve and innovate their processes, methods, and statistical products to better measure an ever-changing world.”⁶

14. The Census Bureau enjoys higher self-response rates than private survey companies in large part because the public generally trusts the Census Bureau to keep its data safe. The Census Bureau makes extensive outreach efforts to assure respondents and other data providers about the Bureau’s commitment to protection of confidential data. The criminal fines and imprisonment penalties that Census Bureau employees would face by unlawfully disclosing respondent information are frequently cited by the Census Bureau in these outreach efforts.⁷
15. This trust in the Census Bureau is particularly important for the decennial census, given the “civic ceremony” aspect of the census, akin to the civic ceremony aspect of elections and voting. The decennial census is an exercise where the nation comes together every ten years, under a strict promise of confidentiality, to provide information to help govern our nation. Were the Census Bureau to expose confidential information, there is no doubt that self-response rates would drop, increasing survey

⁴ National Academies of Sciences, Engineering, and Medicine 2021. Principles and Practices for a Federal Statistical Agency: Seventh Edition. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25885>, p. 37-38.

⁵ Ibid., p.38.

⁶ Ibid., p.4.

⁷ <https://www.census.gov/content/dam/Census/library/factsheets/2019/comm/2020-confidentiality-factsheet.pdf>.

cost across programs by increasing in-person follow up, and decreasing the quality of the census overall.

PRIVACY PROTECTION AT THE CENSUS BUREAU

16. Protecting privacy is at the core of the Census Bureau's mission. Our privacy promise to respondents is key to promoting response to our censuses and surveys. The Census Bureau—at the crux of its dual mandate to publish only statistical summaries and to protect the confidentiality of respondent data—is balancing the preferences of data users and data providers. An optimal choice must account for the preferences of data users and protect the data the American people entrust the Census Bureau with keeping safe.⁸
17. Data collected from the decennial census support a wide array of critical government and societal functions at the federal, state, tribal, and local levels. In addition to apportioning seats in the U.S. House of Representatives and supporting the redistricting of those seats, census data also support the allocation of over \$675 billion in federal

⁸ "Official Statistics at the Crossroads: Data Quality and Access in an Era of Heightened Privacy Risk," *The Survey Statistician*, 2021, Vol. 83, 23-26 (available at [Survey-Statistician-2021-January-N83-03.pdf \(isi-iass.org\)](https://www.isi-iass.org/Survey-Statistician-2021-January-N83-03.pdf)). The paper is based on talks that I gave in 2019 to the Committee on National Statistics and the Joint Statistical Meetings. It summarizes the research in Abowd, J.M. and I. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review*, Vol. 109, No. 1 (January 2019):171-202, DOI:[10.1257/aer.20170627](https://doi.org/10.1257/aer.20170627).

funding each year based on population counts, geography, and demographic characteristics.⁹ Census data also support important public and private sector decision-making at the federal, state, tribal, and local levels, and serve as benchmark statistics for other important surveys and data collections throughout the decade.¹⁰

18. The Census Bureau publishes an enormous number of statistics calculated from its collected data. Following the 2010 Census, for example, the Census Bureau published over 150 billion independent statistics about the characteristics of the 308,745,538 persons in the resident population that were enumerated in the census. To serve their intended governmental and societal uses, the majority of these statistics needed to be published at very fine levels of detail and with geographic precision often down to the individual census tract or block.
19. While it would be quite difficult from any single one of those published statistics to ascertain the identity of any individual census respondent or the contents of that respondent's census response, the volume and detail of information published by the Census Bureau, taken together, pose a serious challenge for protecting the privacy and confidentiality of census data. Combining information from multiple published statistics or tables can make it easy to pick out those individuals in a particular geographic area whose characteristics differ from those of the rest of their neighbors. These individuals, who have unique combinations of the demographic characteristics

⁹ Hotchkiss, M., & Phelan, J. (2017). Uses of Census Bureau data in federal funds distribution: A new design for the 21st century. United States Census Bureau. <https://www2.census.gov/programs-surveys/decennial/2020/program-management/working-papers/Uses-of-Census-Bureau-Data-in-Federal-Funds-Distribution.pdf>.

¹⁰ Sullivan, T. A. (2020). Coming to Our Census: How Social Statistics Underpin Our Democracy (and Republic). *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.c871f9e0>.

reported in statistical summaries, are known as “population uniques” and their records have traditionally been the target of the mechanisms that the Census Bureau uses to protect confidentiality in its data publications.

20. Traditional statistical disclosure limitation methods,¹¹ like those used in 2010 census, cannot defend against modern challenges posed by enormous cloud computing capacity and sophisticated software libraries. That does not mean traditional statistical disclosure limitation methods usually fail – they usually do not fail. But as computer scientists bring their expertise from the field of cryptography to the field of safe data publication, they have exposed significant vulnerabilities in traditional privacy methods. The Census Bureau’s own internal analysis, for example, confirmed that a modern database reconstruction-abetted re-identification attack can reliably match a large number of 2010 census responses to the names of those respondents – a vulnerability that exposed information of *at least* 52 million Americans and potentially up to 179 million Americans.¹² To defend against this known vulnerability, the Census Bureau explored different confidentiality methods that explicitly defend against database reconstruction attacks and concluded that the best tool to protect against this modern attack while also preserving the accuracy and usability of data products comes from the body of scientific work called “differential privacy.”

THE HISTORY OF INNOVATION IN THE DECENNIAL CENSUS

21. The decennial census, known officially as the *Decennial Census of Population and Housing*, is the flagship statistical product of the U.S. Census Bureau. Though the Census

¹¹ The technical field that addresses confidentiality is known as “statistical disclosure limitation.” At the Census Bureau, it is known as “disclosure avoidance.” It is also called “statistical disclosure control” by some statisticians and “privacy-preserving data analysis” by some computer scientists.

¹² See Appendix B for a summary of the Census Bureau’s simulated reconstruction and re-identification attacks.

Bureau conducts hundreds of surveys every year, the once-every-decade enumeration of the population of the United States, mandated by Article I, Section 2 of the U.S. Constitution, is the single largest and most complex data collection regularly conducted by the United States government. Since the very first U.S. census in 1790, the collection, processing, and dissemination of census data have posed unique challenges and have required the Census Bureau to improve its operations every decade.

22. The challenges faced by the Census Bureau have led to remarkable innovations. Herman Hollerith's electric tabulation machine, developed for the 1890 Census, revolutionized the field of data processing and led Hollerith to form the company that eventually became IBM.¹³ To conduct the 1950 Census, the Census Bureau commissioned the development of the first successful civilian digital computer, UNIVAC I.¹⁴ With each passing decade, the Census Bureau develops, tests, and deploys innovations to its statistical methods, field data collection methods, and data processing operations.

23. That spirit of innovation includes the Census Bureau's more recent implementation of cutting-edge privacy protections. Prior to the 1990 Census, the primary mechanism that the Census Bureau employed to protect the confidentiality of individual census responses was to withhold publication of (or "suppress") any table that did not meet certain household, population, or demographic characteristic thresholds. The 1970 Census, for example, suppressed tables reflecting fewer than five households, and would only publish tables of demographic characteristics cross-tabulated by race if

¹³ https://www.census.gov/history/www/census_then_now/notable_alumni/herman_hollerith.html.

¹⁴ https://www.census.gov/history/www/innovations/technology/univac_i.html.

there were at least five individuals in each reported race category.¹⁵ These suppression routines helped to protect privacy by reducing the detail of data published about individuals who were relatively unique within their communities. By the 1990 Census, however, the Census Bureau transitioned away from suppression methodologies for two reasons: first, data users were dissatisfied with missing details caused by suppression and second, the Bureau realized that the suppression routines it had been using were insufficient to fully protect against re-identification.¹⁶

24. For the 1990 Census, the Bureau began using a technique known as noise infusion to safeguard respondent confidentiality. Noise infusion helps to protect the confidentiality of published data by introducing controlled amounts of error or “noise” into the data. The goal of noise infusion is to preserve the overall statistical validity of the resulting data while introducing enough uncertainty that attackers would not have any reasonable degree of certainty that they had isolated data for any particular respondent. The noise infusion used in 1990 was a very simple form of data swapping between paired households in a geographic area with similar attributes, and for small

¹⁵ Zeisset, P. (1978), “Suppression vs. Random Rounding: Disclosure Avoidance Alternatives for the 1980 Census,” <https://www.census.gov/content/dam/Census/library/working-papers/1978/adrm/Suppression%20vs.%20Random%20Rounding%20Disclosure-Avoidance%20Alternatives%20for%20the%201980%20Census.pdf>.

¹⁶ McKenna, L. (2018), “Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing,” <https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf>, p.6.

block groups the Census Bureau replaced the collected characteristics of households with imputed characteristics.¹⁷

25. For the 2000 and 2010 censuses, the Census Bureau began to infuse noise using a more advanced “data swapping” method. The Census Bureau first identified households most vulnerable to re-identification—especially households on smaller-population blocks whose residents had differing demographic characteristics from the remainder of their block. While every non-imputed¹⁸ household record in the Census Edited File (CEF) had a chance of being selected for data swapping, records for more vulnerable households (typically those on low-population blocks) were selected with greater probability. Then, the records for all members of those selected households were exchanged with the records of households in nearby geographic areas that matched on key characteristics. For the 2000 and 2010 censuses, those key matching characteristics were (1) the whole number of persons in the household, and (2) the whole number of persons aged 18 or older in the household. These swapping criteria resulted in the total population and total voting age population for each block being held “invariant”—that is, while noise was added to all remaining characteristics, no noise was added to the block-level total population or block-level voting age population

¹⁷ Ibid., p. 6-7. An “imputed characteristic” is the prediction of a statistical model used in place of a missing characteristic, when used in standard editing procedures, or in place of a collected characteristic, when used for confidentiality protection.

¹⁸ When a respondent household provides only a count of the number of persons living at that address or when the housing unit population count is itself imputed, the Census Bureau imputes all characteristics: sex, age, race, ethnicity, and relationship to others in the household. Such persons are called “whole-person census imputations” in technical documentation. When a household consists entirely of whole-person census imputation records, it is called an “imputed” household. A “non-imputed” household contains at least one person whose characteristics were collected on the census form for the household.

counts.¹⁹ *The selection and application of these particular invariants is not an innate feature of data swapping; invariants are implementation parameters that can be applied to (or removed from) any counted characteristic under any noise infusion methodology.*

THE PRIVACY PROTECTIONS USED FOR THE 2010 CENSUS ARE NO LONGER SUFFICIENT

26. While the Census Bureau's confidentiality methodologies for the 2000 and 2010 censuses were considered sufficient at the time, advances in technology in the years since have reduced the confidentiality protection provided by data swapping.
27. Disclosure avoidance has been a recognized branch of statistics since the 1970s, but it has only been since the late 1990s that it has evolved into a distinct scientific field of study in both statistics and computer science. Prof. Latanya Sweeney's 1997 revelation that she had re-identified then Massachusetts Governor William Weld's medical records in a purportedly "deidentified" public database²⁰ prompted the Census Bureau and many other statistical agencies to re-examine the efficacy of their disclosure avoidance techniques.
28. *Re-identification attacks.* Prior to 2016, disclosure risk assessments usually focused on assessing the vulnerability of microdata releases (data products that contain individual records for all or some of the data subjects deidentified by removing names and addresses), rather than the rules used for aggregated data releases (data compiled and aggregated into tables). Simulated "re-identification attacks" analyze the risk that an external attacker could use individuals' characteristics that are included on a published microdata file (e.g., location, age, and sex) and link those records to a third-

¹⁹ Ibid. p. 8-10.

²⁰ Sweeney, L. (2002). "k-anonymity: a model for protecting privacy." *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5); 557-570, also recounted in Ohm, P. (2009) "Broken promises of privacy: Responding to the surprising failure of anonymization." *UCLA L. Rev.* 57: 1701.

party data source (e.g., commercial data or voter registration lists) that contains those characteristics along with the individuals' names and addresses. The resulting rates of "putative" (suspected) and confirmed linkages show the overall degree of vulnerability of the data. If those linkage rates are deemed too large, then additional disclosure avoidance is necessary to mitigate the disclosure risk.

29. The general problem with relying exclusively on re-identification studies to assess disclosure risk is that they can only provide a "best-case" approximation of the underlying disclosure risk of the data. If a real attacker has access to more sophisticated tools (e.g., optimization algorithms or computing power) or to higher quality external data (e.g., with better age and address information) than the tools or data used in the simulated attack, then the real disclosure risk will be substantially higher than what is estimated via the study. This limitation is particularly vexing for statistical agencies that must rely on a "release and forget" approach to data publication, where disclosure avoidance safeguards must be selected without foreknowledge of the better tools and external data that attackers may have at their disposal after the data are published.

30. Re-identification studies also underestimate the risk from releasing aggregated data. The Census Bureau has long relied on re-identification studies to assess the disclosure risk of its microdata releases, but the majority of Census Bureau data products are aggregated data releases. Over the past decade, aggregated data releases have become increasingly vulnerable to sophisticated "reconstruction attacks" that have emerged as computing power has improved and gotten substantially cheaper.

31. *Reconstruction attacks.* The theory behind a “reconstruction attack” is that the release of *any* statistic calculated from a confidential data source will reveal a potentially trivial, but non-zero, amount of confidential information.²¹ As a consequence, if an attacker has access to enough aggregated data with sufficient detail and precision, then the attacker may be able to leverage information from each statistic in the aggregated data to reconstruct the individual-level records that were used to generate the published tables. This process is known as a “reconstruction attack,” and it adds a new degree of disclosure vulnerability against which statistical agencies must defend. While the statistical and computer science communities have been aware of this vulnerability since 2003, only over the last few years have computing power and the sophisticated numerical optimization software necessary to perform these types of reconstructions advanced enough to permit reconstruction attacks at any significant scale.
32. The risk of reconstruction and re-identification attacks is real and substantiated. The Census Bureau has been approached by Prof. Sweeney and others who claim that they have identified specific vulnerabilities in our standard disclosure avoidance methodologies.²² The vulnerabilities in the disclosure avoidance protections for the Census Bureau’s Survey of Income and Program Participation (SIPP) identified by Prof. Sweeney led the Census Bureau to immediately implement permanent changes to the

²¹ Dinur, I. and Nissim, K. (2003) “Revealing Information while Preserving Privacy” PODS, June 9-12, San Diego, CA. <https://doi.org/10.1145/773153.773173>.

²² McKenna, L. (2019b). “U.S. Census Bureau Reidentification Studies,” available at <https://www.census.gov/library/working-papers/2019/adrm/2019-04-ReidentificationStudies.html>.

disclosure avoidance rules used for SIPP data, including increased noise infusion and delayed reporting of survey participants' major life events.²³

33. Statistical releases do not all need to be of the same type, or contain the same data fields, to enable re-identification by reconstruction. For example, a 2015 interagency report published by the National Institute of Standards and Technology (NIST) written by my colleague Simson Garfinkel provided examples of using disparate data sets to reconstruct hidden underlying data.²⁴ Some of these examples are quoted here:

34. "*The Netflix Prize*: Narayanan and Shmatikov showed in 2008 that in many cases the set of movies that a person had watched could be used as an identifier.²⁵ Netflix had released a dataset of movies that some of its customers had watched and ranked as part of its "Netflix Prize" competition. Although there was [sic] no direct identifiers in the dataset, the researchers showed that a set of movies watched (especially less popular films, such as cult classics and foreign films) could frequently be used to match a user profile from the Netflix dataset to a single user profile in the Internet Movie Data Base (IMDB), which had not been de-identified and included user names, many of which were real names. The threat scenario is that by rating a few movies on IMDB, a person might inadvertently reveal *all* of the movies that they had watched, since the person's IMDB profile could be linked with the Netflix Prize data."²⁶ (emphasis in original)

²³ McKenna, L. (2019b). p. 2-3.

²⁴ Garfinkel, S. (2015) "De-Identification of Personal Information," National Institute of Standards and Technology, available at <http://dx.doi.org/10.6028/NIST.IR.8053> at 26-27.

²⁵ Narayanan, A. and Shmatikov V. "Robust De-anonymization of Large Sparse Datasets," *IEEE Symposium on Security and Privacy* (2008): 111-125.

²⁶ Garfinkel, S. (2015), p. 26-27.

35. “*Credit Card Transactions*: Working with a collection of de-identified credit card transactions from a sample of 1.1 million people from an unnamed country, Montjoye *et al.* showed that four distinct points in space and time were sufficient to specify uniquely 90% of the individuals in their sample.²⁷ Lowering the geographical resolution and binning transaction values (*e.g.*, reporting a purchase of \$14.86 as between \$10.00 and \$19.99) increased the number of points required.”²⁸
36. “*Mobility Traces*: Montjoye *et al.* showed that people and vehicles could be identified by their “mobility traces” (a record of locations and times that the person or vehicle visited). In their study, trace data from a sample of 1.5 million individuals was processed, with time values being generalized to the hour and spatial data generalized to the resolution provided by a cell phone system (typically 10-20 city blocks).²⁹ The researchers found that four randomly chosen observations of an individual putting them at a specific place and time was sufficient to uniquely identify 95% of the data subjects.³⁰ Space/time points for individuals can be collected from a variety of sources, including purchases with a credit card, a photograph, or Internet usage. A similar study performed by Ma *et al.* found that 30%-50% of individuals could be identified with 10 pieces of side information.³¹ The threat scenario is that a person who

²⁷ Montjoye, Y-A. et al. “Unique in the shopping mall: On the reidentifiability of credit card metadata,” *Science*, 30 (January 2015) Vol 347, Issue 6221.

²⁸ Garfinkel, S. (2015), p. 27.

²⁹ De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 3(1).

³⁰ *Ibid.*, p. 1-5.

³¹ C. Y. T. Ma, D. K. Y. Yau, N. K. Yip and N. S. V. Rao (2013) "Privacy Vulnerability of Published Anonymous Mobility Traces," in *IEEE/ACM Transactions on Networking*, vol. 21, no. 3, pp. 720-733, June 2013, doi: 10.1109/TNET.2012.2208983.

revealed five place/time pairs (perhaps by sending email from work and home at four times over the course of a month) would make it possible for an attacker to identify his or her entire mobility trace in a publicly released dataset. As above, the attacker would need to know that the target was in the data.”³²

37. The same general principles apply to census data. The difference between census data and the examples above is that census data can be combined in vastly more ways with other information because all the tables published from census data share basic standardized identifiers including location, age, sex, race, ethnicity, and marital status. Even if each of these identifiers is not included in every table, their use and combinations across many different tables creates the disclosure risk. The Census Bureau understood this emerging risk even before the 2010 Census. As field collection for the 2010 Census was first beginning, the Census Bureau had already flagged the heightened disclosure risk of releasing detailed block level population data, even with the 2010 Census swapping mechanism in place.³³ After tracking this growing risk of reconstruction and re-identification attacks for several years, the Census Bureau decided in 2015 to establish a new team to comprehensively evaluate the Census Bureau’s disclosure avoidance methods to determine if they were sufficient to protect against these disclosure risks.³⁴

³² Garfinkel, S. (2015), p. 27-28.

³³ During a January 2010 meeting of the Census Bureau’s Data Stewardship Executive Policy (DSEP) Committee, the chair of the Disclosure Review Board voiced her concerns about the 2010 Census swapping mechanism’s ability to adequately protect future censuses, noting specifically the challenge posed by “continuing to release data at the block level, as block populations continue to decrease (e.g., 40% of blocks in North Dakota have only 1 household in them)” Based on this warning, DSEP decided that “the problem of block population size and disclosure avoidance is real, and that it deserves attention in the context of 2020 planning.” DSEP Meeting Record, January 14, 2010. See Appendix C.

³⁴ DSEP Meeting Record, February 5, 2015. See Appendix D.

2010 CENSUS SIMULATED RECONSTRUCTION-ABETTED RE-IDENTIFICATION ATTACK

38. The results from the Census Bureau's 2016-2019 research program on simulated reconstruction-abetted re-identification attack were conclusive, indisputable, and alarming. Appendix B, attached to this declaration, provides an overview of that simulation and the results. The bottom line is that our simulated attack showed that a conservative attack scenario using just 6 billion of the over 150 billion statistics released in 2010 would allow an attacker to accurately re-identify *at least* 52 million 2010 Census respondents (17% of the population) and the attacker would have a high degree of confidence in their results with minimal additional verification or field work. In a more pessimistic scenario, an attacker with access to higher quality commercial name and address data than those used in our simulated attack could accurately re-identify around 179 million Americans or around 58% of the population.
39. Emerging attack scenarios and our own internal simulated attacks show that were the Census Bureau to use the disclosure avoidance mechanism implemented for the 2010 Census again for the 2020 Census, the results would be vulnerable to reconstruction and re-identification attacks because of the parameters of the swapping mechanism's 2010 implementation: an overall insufficient level of noise, the invariants preserved without noise, and the geographic and demographic detail of the published summary data. The Census Bureau can no longer rely on the swapping implementation used in 2010 if it is to meet its obligations to protect respondent confidentiality under 13 U.S. Code §§ 8(b) & 9. Protecting against new technology-enabled re-identification attacks, while maintaining the high quality of the decennial census data products, requires the implementation of a disclosure avoidance mechanism that is better able to protect against these new, sophisticated vectors of attack.

DISCLOSURE AVOIDANCE OPTIONS CONSIDERED FOR THE 2020 CENSUS

40. Faced with this compelling mathematical and empirical evidence of the inherent vulnerability of the 2010 Census swapping mechanism to protect against reconstruction-abetted re-identification attacks, the Census Bureau began exploring the available data protection strategies that it could employ for the 2020 Census. The three methods the Census considered were *Enhanced Data Swapping*, *Suppression*, and *Differential Privacy*.
41. The Census Bureau decided that differential privacy was the best tool after analyzing the various options through the lens of economics. Efficiently protecting privacy can be viewed as an economic problem because it involves the allocation of a scarce resource—confidential information—between two competing uses: public data products and privacy protection. If we produce more accuracy, we will have less privacy, and vice versa. And just like in the classic economic example of the trade-off between producing guns and butter, the tradeoff between privacy and accuracy can be analyzed with a production possibility curve. Our empirical analysis showed that differential privacy offered the most efficient trade-off between privacy and accuracy—our calculations showed that the efficiency of differential privacy dominated traditional methods.³⁵ In other words, regardless of the level of desired confidentiality, differential privacy will always produce more accurate data than the alternative traditional methods considered by the Census Bureau.
42. *Enhanced Data Swapping*. Enhancing the data swapping mechanism used for the 2010 Census in a manner sufficient to protect against emerging threats like reconstruction

³⁵ See Abowd, J. M., & Schmutte, I. M. (2019). An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109(1), 171-202.

attacks would have a significant, detrimental impact on data quality. With an estimated 57% of the population³⁶ known to be unique at the block level, a swapping mechanism that targets vulnerable households for swapping would require significantly higher rates of swapping than were used in 2010 to protect against a reconstruction attack. Implementing swapping in 2020 would also require abandoning the total population and voting-age population invariants that were used in 2010. There are two technical reasons for this. First, at swap rates sufficient to counter the reconstruction of microdata accurate enough to enable large-scale reidentification, it is impossible to find enough paired households with the same number of persons and adults without searching well outside the neighborhood of the original household. Finding swap pairs was a challenge for some states even at the 2010 swap rate. Second, holding the total and adult populations invariant gives the attacker a huge reconstruction advantage—exact record counts in each block for persons and adults. This advantage vastly improves the accuracy of the reconstructed data. Even a small amount of uncertainty about the block location of an individual greatly expands the variability in the reconstructed microdata effectively reducing the chances of a correct linkage in a re-identification attack. If a block is known to contain exactly seven persons in the confidential data, then every feasible reconstructed version of those data will have exactly seven records in that block, meaning that the block identifier will be correct on every record of every feasible reconstructed database. But if the block population is reported with some random fluctuation around seven, then only by chance will the

³⁶ Fifty-seven percent of the 308,745,538 person records in the confidential 2010 Census Edited File, the definitive source for all 2010 Census tabulations, were unique on their block location, sex, age (in years), race (any combination of the 6 OMB-approved race categories, 63 possibilities in all) and Hispanic/Latino ethnicity. This previously confidential statistic was approved for publication with DRB clearance number CBDRB-FY21-DSEP-003.

block identifier be correct in the reconstructed data. Compound this effect over 8,000,000 blocks and the number of feasible reconstructions explodes exponentially. This is what provides the protection against re-identification from the reconstructed data.³⁷ Internal experiments also confirmed that increasing the swap rate from the level used in 2010 and removing the invariants on block-level population counts (to permit the increased level of swapping and protect against reconstruction attacks) would render the resulting data unusable for most data users.

43. *Suppression*. While the Census Bureau could use suppression to protect from a reconstruction attack, the resulting data would be only available at a very high level of generality. Today's data users, including redistricters, rely on detailed block and tract-level data, which would not be available for many areas if the Census were to return to suppression to protect against modern attacks.
44. *Differential Privacy*. Differential privacy, first developed in 2006, is a framework for quantifying the precise disclosure risk associated with each incremental release from a confidential data source.³⁸ In turn, this allows an agency like the Census Bureau to quantify the precise amount of statistical noise required to protect privacy. This precision allows the Census to calibrate and allocate precise amounts of statistical noise in a way that protects privacy while maintaining the overall statistical validity of the data.

³⁷ Garfinkel, S., Abowd, J. M., & Martindale, C. (2018). Understanding Database Reconstruction Attacks on Public Data: These attacks on statistical databases are no longer a theoretical danger. *Queue*, 16(5), 28-53.

³⁸ Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* (pp. 265-284). Springer, Berlin, Heidelberg.

45. The Census Bureau first began using differential privacy to protect its statistical data products in 2008, with the launch of its [OnTheMap](#) tool for employee commuting statistics and its heavily used extension [OnTheMap for Emergency Management](#). In the years since, the Census Bureau has also successfully used differential privacy in a number of other innovative statistical products, such as the Post-Secondary Employment Outcomes and Veteran Employment Outcomes products. Differential privacy is also being used by many of the major technology firms, including Apple³⁹, Google,⁴⁰ Microsoft,⁴¹ and Uber.⁴² Other statistical agencies, such as the Statistics of Income Division of the Internal Revenue Service, have also begun implementing differential privacy.⁴³ Internationally, the Australian Bureau of Statistics,⁴⁴ the Office of National

³⁹Differential Privacy Team. (2017). “Learning with Privacy at Scale.” Apple Machine Learning Journal, 1(8).

⁴⁰Erlingsson, U., V. Pihur, and A. Korolova. (2014). “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response.” Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS ’14, 1054–1067.

⁴¹ Ding, B., J. Kulkarni, and S. Yekhanin. (2017). “Collecting Telemetry Data Privately.” Advances in Neural Information Processing Systems 30.

⁴² Near, J. (2018) “Differential Privacy at Scale: Uber and Berkeley Collaboration,” Enigma 2018 (January) USENIX Assoc. <https://www.usenix.org/node/208168>.

⁴³ Bowen, C. et al. (2020) “A Synthetic Supplemental Public-Use File of Low-Income Information Return Data: Methodology, Utility, and Privacy Implications,” (July) Tax Policy Center, The Brookings and Urban Institutes. https://www.urban.org/sites/default/files/publication/102547/a-synthetic-supplemental-public-use-file-of-low-income-information-return-data_2.pdf.

⁴⁴ Australian Bureau of Statistics, (2019) “Protecting the Confidentiality of Providers,” January 2019, 1504.0 - Methodological News, <https://www.abs.gov.au/ausstats/abs@.nsf/Previousproducts/1504.0Main%20Features9999Jan%202019?opendocument&tabname=Summary&prodno=1504.0&issue=Jan%202019&num=&view=>, accessed on March 31, 2021.

Statistics in the United Kingdom,⁴⁵ and Statistics Canada⁴⁶ explicitly recognize the threat from combining multiple statistical tabulations to re-identify respondent information and recommend output noise infusion systems, including differential privacy.

46. Faced with the alarming results of the simulated reconstruction attack, which indicated that the established swapping mechanism resulted in far less disclosure protection than it was intended to provide, and considering the available alternatives, the Census Bureau's Data Stewardship Executive Policy Committee (DSEP)⁴⁷ determined that the Census Bureau should proceed with the deployment and testing of differential privacy for use in the 2020 Census given its obligations to produce high quality statistics from the decennial census while also protecting the confidentiality of respondents' census records under 13 U.S. Code §§ 8(b) & 9.⁴⁸

⁴⁵ United Kingdom Office for National Statistics, (2021) "Policy on Protecting Confidentiality in Tables of Birth and Death Statistics," <https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/policyonprotectingconfidentialityintablesbirthanddeathstatistics#annex-a-understanding-the-legal-and-policy-framework>, accessed on March 31, 2021.

⁴⁶ Statistics Canada, (2021) "A Brief Survey of Privacy Preserving Technologies," March 2021, *Data Science Network for the Federal Public Service*, <https://www.statcan.gc.ca/eng/data-science/network/privacy-preserving>, accessed on March 31, 2021.

⁴⁷ The Data Stewardship Executive Policy Committee (DSEP) is a committee chaired by the Deputy Director/Chief Operating Officer and composed of career senior executives with expertise in confidentiality practice, the uses of Census Bureau data, and policy. DSEP is the parent organization for the Disclosure Review Board (DRB), which reviews and approves individual data releases to ensure that no confidential data is released.

⁴⁸ On May 10-11, 2017 DSEP decided that "any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB" meaning that all 2020 Census publications would be subject to differential privacy. See Appendices E and F. On February 15, 2018 DSEP suspended publication of "all proposed tables in Summary File 1 and Summary File 2 for the 2020 Census at the block, block-group, tract, and county level except for the PL94-171 tables, as announced in Federal Register Notice 170824806-7806-01..." acknowledging that "...these data in many

47. The best disclosure avoidance option that offers a solution capable of addressing the new risks of reconstruction-abetted re-identification attacks, while preserving the fitness-for-use of the resulting data for the important governmental and societal uses of census data, is differential privacy. I have summarized here what I consider to be the most important reasons that the Census Bureau decided to adopt differential privacy.
48. **Disclosure avoidance must be proactive.** The fundamental objective of disclosure avoidance protections is to proactively prevent disclosures. Just like corporations are not expected to wait until they have suffered a major data breach before upgrading their IT security systems to protect against known threats, statistical agencies should not wait until they suffer a confirmed breach before improving their disclosure avoidance protections to account for known threats. The expectation, for both IT security and disclosure avoidance, is to remain vigilant about emerging threats and risks, and to take appropriate action *before* those risks lead to a breach.
49. **The privacy risk landscape has fundamentally changed since 2010.** Traditional methods of assessing disclosure risk rely on knowing what tools and resources an attacker might leverage to undermine confidentiality protections. These tools, however, are ever evolving. Over the last decade, technological advances have made powerful cloud computing environments, with sophisticated optimization algorithms

cases were accurate to a level that was not supported by the actual uses of those data, and such an approach is simply untenable in a formally private system.” DSEP further decided that “SF1 and SF2 will be rebuilt based on use cases.” See Appendix G. In parallel with these decisions by DSEP, the disclosure risks identified by the preliminary results of the simulated reconstruction attack also led to this issue being added to the Census Bureau’s risk management portfolio. On April 17, 2017 the risk of reconstruction attacks was proposed for inclusion in the Research and Methodology Directorate’s risk registry. On September 12, 2017 it was escalated and included on the Enterprise-level Risk register. Finally, on January 30, 2018, it was further escalated to the Enterprise-level Issue register, with the development and use of the 2020 Census Disclosure Avoidance System as an identified resolution action to be taken. .

capable of performing large-scale attacks, cheap and easily available. While these tools were not yet a viable attack model in 2010, they certainly represent a credible threat today.⁴⁹

50. **Internal research has conclusively proven the fundamental vulnerabilities of the 2010 swapping methodology.** The Census Bureau has performed extensive empirical analysis of the disclosure risk inherent to the 2010 Census swapping methodology as detailed in Appendix B. No technique can produce usable data with absolutely zero risk of re-identification, but the re-identification rates from our internal experiments on the 2010 Census swapping methodology are orders of magnitude higher than what they were intended to be. The privacy threat landscape has evolved over the last decade and compels the Census Bureau to adapt its protections accordingly.
51. **The Census Bureau determined that differential privacy was the only method that could adequately protect the data while preserving the value of census data products.** When our internal research demonstrated the vulnerabilities of the swapping mechanism used for the 2010 Census, we considered a range of options for the 2020 Census. The three leading options were differential privacy, an enhanced version of data swapping, and a return to whole-table suppression. But to achieve the necessary level of privacy protection, both enhanced data swapping and suppression had severely deleterious effects on data quality and availability. With its enhanced privacy protections and precision control over the tuning of privacy/accuracy tradeoff, the Census Bureau determined that differential privacy was the only viable solution for the 2020 Census.

⁴⁹ DSEP drew this conclusion from the simulated reconstruction-abetted re-identification attack in Appendix B. The Office of National Statistics reached the same conclusion in its 2018 “Privacy and data confidentiality methods: a Data and Analysis Method Review (DAMR)” at [Privacy and data confidentiality methods: a Data and Analysis Method Review \(DAMR\) – GSS \(civilservice.gov.uk\)](#) (cited on April 10, 2021).

52. Differential privacy can be fine-tuned to strike a balance between privacy and accuracy. DSEP made the preliminary decision to pursue differential privacy on May 10-11, 2017. Since that decision was announced, the Census Bureau has worked extensively with our advisory committees, federal agency partners, American Indian and Alaska Native tribal leaders, the Committee on National Statistics, professional associations, data user groups, and many others at the national, state, and local levels to understand how they use decennial census data and to ensure that our implementation of differential privacy will preserve the value of the decennial census as a national resource. The Census also released sets of demonstrative data to allow the public and end-users to provide feedback that allowed us to fine-tune and tweak how we will ultimately implement differential privacy.⁵⁰

53. The need to modernize our privacy protections has been confirmed by external experts. The Census Bureau's ongoing partnerships with scientific and academic experts from around the country helped us conduct the internal evaluation of the disclosure risk of the 2010 Census swapping methodology and confirmed the need to modernize our privacy protections. To supplement this ongoing work and to get external expert confirmation of the conclusions that we have drawn from it, the Census Bureau also commissioned an independent expert review by JASON, an independent group of elite scientists that advise the federal government on science and technology. The JASON report confirmed our findings regarding the re-identification risk inherent to the 2010 Census swapping methodology.⁵¹

⁵⁰ U.S. Census Bureau "Developing the DAS: Demonstration Data and Progress Metrics" <https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-development.html>.

⁵¹ JASON (2020). "Formal Privacy Methods for the 2020 Census" JASON Report JSR-19-2F. <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/privacy-methods-2020-census.pdf>.

54. Differential Privacy can produce highly accurate data. One key benefit of differential privacy is the ability to fine-tune privacy and accuracy. The next iteration of demonstration data will establish that differential privacy protections can produce extremely accurate redistricting data. While the full April 2021 Demonstration Data Product⁵² and supporting metrics will be released by April 30, 2021, I can provide a high-level summary of key metrics:⁵³

- Total populations for counties have an average error of +/- 5 persons (reflecting a mean absolute percent error of 0.04% of the counties' population) as noise from differential privacy.⁵⁴ This is extremely accurate considering that if we simulate the errors in census counts as estimates of the true population, then the average county-level estimation uncertainty of the census is +/- 960 persons (averaging 1.6% of the county census counts).⁵⁵

⁵² The April 2021 demonstration data uses a global privacy-loss budget of 10.3 with a very substantial proportion allocated to detailed race and ethnicity statistics at the block and block group levels.

⁵³ Statistics for the April 2021 Demonstration Data Product are preliminary, based on the internal research version. The production version will be used for the detailed summary statistics when they are posted on census.gov.

⁵⁴ The statistics are the mean absolute error and the mean absolute percentage error in county population comparing the April 2021 Demonstration Data Product to the data released in the 2010 Summary File 1.

⁵⁵ The inherent error in the census counts as estimates of the true population can be simulated using data-defined person and correct-enumeration rates from coverage measurement estimates, in this case from the most recent decennial census in 2010. (See Mule, T. "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States", Report G-10, g01.pdf (census.gov). Table 3, in particular.) An alternative modeling perspective simulates the natural variation of census population estimates using the natural variation in census estimates due to erroneous enumerations and other sources of error inherent in the Census. For county populations

- At the block level the differentially private data have an average population error of ± 3 persons, which includes both housing unit and group quarters populations. Compare that with the simulated error inherent in the census which puts the average error uncertainty of block population counts at ± 6 people.⁵⁶

55. **The April 2021 demonstration data show no meaningful bias in the statistics for racial and ethnic minorities** even in very small population geographies like Federal American Indian Reservations. The data permit assessment of the largest OMB-designated race and ethnicity group in each geography – the classification used by the Department of Justice for Voting Rights Act scrutiny – with a precision of 99.5% confidence in variations of ± 5 percentage points for off-spine geographies as small as 500 persons, approximately the minimum voting district size in the redistricting plans that the Department of Justice provided as examples.

56. **The accuracy of differential privacy increases at higher levels of geography, even for arbitrary geographic areas like Congressional and legislative districts.** The Census Bureau designed its implementation of differential privacy to increase accuracy

this natural variation is about ± 120 persons (0.3% of population), also based on coverage data from the 2010 Census. As with all simulation estimates, there is sensitivity to the assumptions. The reported statistics are the mean absolute error and the mean absolute percentage error. Differentially private statistics include both the housing unit and group quarters populations. Simulations exclude the group quarters population because there are no coverage estimates for that group.

⁵⁶ The simulation of the natural variation of census block-level populations is ± 1.5 persons, which excludes the group quarters population because there are no coverage estimates for that group. As with all simulation estimates, there is sensitivity to the assumptions. The reported statistics are the mean absolute errors. Mean absolute percentage errors are not useful statistics for block populations because more than 2,000,000 blocks with positive housing units have populations between 0 and 9. Differentially private statistics include both the housing unit and group quarters populations. Simulations exclude the group quarters population because there are no coverage estimates for that group.

as blocks are aggregated into larger geographic areas like neighborhoods, voting districts, towns, and other places. Rather than infusing noise at the block level and aggregating upwards, which would cause error to compound at larger geographic levels, the Disclosure Avoidance System's TopDown Algorithm (TDA) takes the opposite approach. Starting at the national level, the algorithm establishes very precise (but still privacy-protected) tabulations for all characteristics at the national level, then works its way down the geographic hierarchy, ensuring that all of the geographic entities at each level (e.g., the Census tracts within a county) add up precisely to the established characteristics of the level above (e.g., the county). This approach limits the distortions that can arise from noise infusion and ensures the reliability of statistics as the underlying size of the population increases. Plaintiffs argue that "the November 2020 demonstration data also skewed the 2010 tabulations enough to create a population deviation in Alabama's Congressional districts on a level that courts have found in other contexts to violate voters' equal population rights," with districts losing up to 73 individuals or gaining 206 individuals over reported values. While this may have been true for the November 2020 Demonstration Data Product, this is not true for the Demonstration Data Product that will be produced by the end of April. In the April 2021 Demonstration Data Product, Congressional districts as drawn in 2010 have a mean absolute percentage error of 0.06%. If the Congressional districts had been drawn using the April 2021 Demonstration Data Product, their statistical composition for the purposes of Voting Rights Act scrutiny would not be affected. Even for state legislative districts, which had average sizes of 159,000 (upper chambers) and 64,000 (lower chamber), the mean absolute percentage errors are 0.09% (upper chambers) and 0.16% (lower chambers), respectively. Such errors are trivial and imply that the difference between districts drawn from the April 2021 Demonstration Data Product and those drawn from the original 2010 P.L. 94-171 Redistricting Data Summary File would be statistically and practically imperceptible. *Most importantly*

for the redistricting use case, the TDA, when properly tuned, ensures that redistricters can remain confident in the accuracy of the population counts and demographic characteristics of the voting districts they draw, despite the noise in the individual building blocks.

IMPLEMENTING DIFFERENTIAL PRIVACY FOR THE 2020 CENSUS

57. Census announced that it planned to use Differential Privacy for the 2020 Census in a few different venues: (1) August 3, 2018, 2020 Census Program Management Review; (2) December 6, [2018, Census Scientific Advisory Committee Meeting](#); and (3) [May 2, 2019, Census National Advisory Committee meeting](#).
58. The Bureau has engaged in a years-long campaign to educate the user community and solicit their views about how differential privacy should be implemented. Census Bureau staff have made hundreds of public presentations, held dozens of webinars, held formal consultations with American Indian and Alaska Native tribal leaders, created an extensive website with plain English blog posts, and conducted regular outreach with dozens of stakeholder groups. We have made presentations to our scientific advisory committees and provided substantial information to oversight entities such as the Government Accountability Office and the Office of the Inspector General.
59. Part of the Bureau's effort to inform the public and solicit feedback involved releasing a series of Demonstration Data Products. There are many different ways to implement differentially private disclosure avoidance mechanisms, and the design and parameters of these mechanisms can substantially impact the fitness-for-use of the resulting data. The Census Bureau's TopDown Algorithm (TDA) was specifically designed to address the reconstruction-abetted re-identification vulnerability risks, while allowing the Bureau to tune the accuracy of the statistics to ensure fitness-for-use.

60. To date, the Census Bureau has released four sets of Demonstration Data Products (in October 2019, May 2020, September 2020, and November 2020). The Census Bureau has received substantial, actionable feedback after each release that has contributed to the system's design and optimization.
61. All four of these demonstration products used a lower privacy-loss budget than we anticipate using for the final 2020 Census data – that is, these demonstration data were purposefully “tuned” to privacy and not “tuned” for producing highly accurate re-districting data. We held the privacy-loss budget roughly the same across these four releases to allow us to compare effects of incremental improvements in the system. After each release, these demonstration files enabled data users to help the Census Bureau identify areas where the algorithm needed to be tuned to meet their specific use cases. While the Census Bureau has not yet set the final privacy-loss budget, we have been clear that all the demonstration data released to date have used a lower privacy-loss budget (more privacy, less accuracy) than will be selected for the final production run of the redistricting data.⁵⁷
62. This degree of transparency into the design and implementation of a disclosure avoidance methodology is unprecedented in the federal government. The Census Bureau has submitted its differential privacy mechanisms, programming code, and system architecture to thorough outside peer review. We have also committed to publicly releasing the entire production code base and full suite of implementation settings and parameters. Many traditional disclosure avoidance methods, most notably swapping techniques, must be implemented in a “black box.” Implementation parameters for these legacy disclosure avoidance methods, especially swapping rates, are often

⁵⁷ Most recently on February 23, 2021 in [The Road Ahead: Upcoming Disclosure Avoidance System Milestones \(govdelivery.com\)](https://www.govdelivery.com/topics/census).

some of the most tightly guarded secrets that the Census Bureau protects. But differential privacy does not rely on the obfuscation of its implementation as a means of protecting the data. The Census Bureau's transparency will allow any interested party to review exactly how the algorithm was applied to the 2020 Census data, and to independently verify that there was no improper or partisan manipulation of the data.

INVARIANTS ARE NOT REQUIRED FOR ACCURACY.

63. Invariants — or data held constant when applying differential privacy — introduce privacy risks and are not necessary to ensure accuracy. Invariants were not well understood either theoretically or empirically in 2016 when the Census Bureau began its research on differential privacy for decennial census data, but we now understand that invariants defeat the privacy protections and must be limited in order to protect the integrity of the system as a whole. Unlike traditional approaches to disclosure avoidance, differentially private noise infusion offers quantifiable and provable privacy guarantees. These guarantees, reflected in the global privacy-loss budget and its allocation to each statistic, serve as a promise to data subjects that there is an inviolable upper bound to the risk that an attacker can learn or infer something about those data subjects through publicly released data products. While that upper bound is ultimately a policy decision, and may be low or high depending on the balancing of the countervailing obligations to produce accurate data and to protect respondent confidentiality, the level of the global privacy-loss budget is central to the ability of the approach to protect the data. Invariants are, by their very nature, the equivalent of assigning infinite privacy-loss budget to particular statistics, which fundamentally violates the central promise of differentially private solutions to controlling disclosure risk. By excluding the accuracy of invariant data elements from the control of the privacy-loss budget, invariants exclude the disclosure risk and potential inferences that can be drawn from those data elements from the formal privacy guarantees. Thus,

instead of being able to promise data subjects that the publication of data products will limit an attacker to being able to infer, at most, a certain amount about them (with that amount being determined by the size of the privacy-loss budget and its allocation to each characteristic), the inclusion of one or more invariants fundamentally excludes attacker inferences about the invariant characteristic(s) from the very nature of that promise. The qualifications and exclusions to the privacy guarantee weaken the strength of the approach and make communicating the resulting level of protection substantially more difficult. This is the reason that DSEP removed the block-level invariant on population and voting-age population. Below the state level, DSEP only authorized block-level invariants that were necessary to conduct the field operations of the 2020 Census: housing unit address counts, and occupied group quarters address counts and types. As noted above, if the block population is reported with some random fluctuation around the confidential value, then only by chance will the block identifier be correct in any potential reconstructed microdata. Compound this effect over 8,000,000 blocks and the number of feasible reconstructions explodes exponentially. This is what provides the protection against re-identification from the reconstructed data.

64. Invariants are not required to improve the accuracy of any statistic processed by differential privacy. Assigning sufficiently high (but not infinite) privacy-loss budget to any statistic can ensure perfect accuracy for that statistic while still allowing the resulting privacy-loss to be communicated in the privacy guarantee. For example, the state-level population of the American Indian and Alaska Native tribal areas has been given sufficient privacy-loss budget to ensure that those populations are presented accurate to the number of persons in the units column; the mean absolute error is 1 person, essentially invariant and the same precision as the state populations themselves. But this solution still requires balancing accuracy and privacy-loss overall. All characteristics cannot have large privacy-loss budget allocations at every geographic

level. If they did, the published tables would be exact images of the confidential data and subject to the same vulnerability as the 2010 Census.

65. The forthcoming April 2021 Demonstration Data Product illustrates this tradeoff. These new demonstration data use a global privacy-loss budget for persons of 10.3, which is much larger than the 4.0 budget used in the earlier releases but is still allocated in a manner that provides a level of protection for every census record and every published characteristic. The April 2021 demonstration data also fully satisfy a tightly specified set of accuracy criteria specialized to the redistricting use case. Specifically, populations, voting-age populations, and the proportion of the largest OMB-designated race and ethnicity groups are all reliable for redistricting and Voting Rights Act scrutiny in arbitrary contiguous block aggregates for both on-spine and off-spine political and legal entities. Because new districts cannot be drawn before the 2020 P.L. 94-171 Redistricting Data Summary File is released, counties, block groups, minor civil divisions, incorporated places, and Census-designated places were all used as on- and off-spine geographic entities for tuning purposes.
66. In the April 2021 Demonstration Data Product, all the targeted small population statistics for race and ethnic groups are far more accurate than in previous demonstration data products, even though no additional invariants were used. The gain in accuracy is entirely due to dedicating more of the privacy-loss budget to the block- and block group-level statistical tables and carefully specifying the differentially private measurements to target the OMB-designated race and ethnicity groups. Biases in the tribal areas' race and ethnicity data were also greatly reduced.
67. The Census Bureau has received substantial feedback from our data user community highlighting distortions that were present in the early versions of our demonstration data, particularly in the version released in October 2019. Based on that feedback, the Census Bureau has identified and corrected the algorithmic sources of those distortions. As these measures of accuracy and bias show, any residual impact of the types

of systematic bias observed in the early demonstration data will be negligible and well within the normal variance and total error typical for a census.

PROCESS AND TIMELINE MOVING FORWARD

68. The operational delays caused by the global COVID-19 pandemic, and the resulting processing schedule changes for production of the redistricting data product shifted the milestone dates for all the systems necessary to produce the data. While the 2020 Census Disclosure Avoidance System is fully operational, and has already passed the Test Readiness Review (TRR) and Production Readiness Review (PRR) milestones on schedule, we have taken advantage of the additional time before the May 20, 2021 Operational Readiness Review (ORR) to perform additional optimization and testing of the system, and to engage in another round of data user evaluation and feedback.
69. The Census Bureau will release another demonstration product by April 30, 2021 using a higher privacy-loss budget (more accuracy) that better approximates the final privacy-loss budget that will likely be selected for the redistricting data product. These new demonstration data will also reflect system design changes that have been made since the last demonstration data release, along with tuning and optimization of the system that have been done specifically to prioritize population count accuracy and the ability to identify majority-minority districts.⁵⁸ The new release will give users yet another opportunity to let the Census know specifically where the data are (or are not yet) sufficiently accurate to meet their requirements.
70. On March 25, 2021, DSEP approved the privacy-loss budget to be used for the next demonstration product. This privacy-loss budget reflects empirical analysis of over

⁵⁸ Users will be able to see the difference between algorithmic improvements and greater privacy-loss budget. At the same time as the main April 2021 Demonstration Data Product is released, the Census Bureau will also release demonstration data using exactly the same software implementation but setting the global privacy-loss budget to 4.0 for persons, as it was in the four previous demonstration data products.

600 full-scale runs of the Disclosure Avoidance System using 2010 Census data. The Census evaluated these experimental runs using accuracy and fitness-for-use criteria for the redistricting use case informed by the extensive feedback we have received from the redistricting community and the Civil Rights Division at the U.S. Department of Justice. Based on this feedback, the privacy-loss budget for the final demonstration product is set to ensure the accuracy of racial demographics for voting districts as small as 500 individuals. With this tuning, the proportion of the largest racial group within even those small state/local voting districts of 500 individuals will be accurate to within five percentage points of the enumerated value at least 95% of the time. As voting district population size increases to any sort of reasonably anticipated legislative district, the error will be miniscule. For example, Congressional and state legislature districts will have significantly higher accuracy for population counts and voting age population counts.

71. Following the release of the new demonstration data, data users and stakeholders will have about a month to submit additional feedback on their analysis and assessment of these data, before DSEP, in early June 2021, sets the privacy-loss budget and system parameters for the production run of the redistricting data product.
72. The production run for creating the Microdata Detail File (the internal name for the file that contains the privacy-protected data) is scheduled to occur between June 26 and July 18, 2021. This roughly three-week period is similar to the period required to implement disclosure avoidance in prior censuses and is not the cause of the delay in the delivery of the redistricting data.
73. As discussed in more detail below, any court-ordered change in the Census Bureau's implementation of disclosure avoidance would add significant time to this schedule.

BRYAN AND BARBER DECLARATIONS

74. Although I cannot set out all my observations and disagreements with the declarations of Dr. Michael Barber and Mr. Thomas Bryan in this declaration, I want to identify some key areas of dispute.
75. Dr. Barber's expert report does not adequately account for the fact that the Census Bureau's demonstration data products had a privacy-loss budget significantly lower than the expected budget that will be set for the 2020 Census. As I explained above, we purposefully set the budget lower than ones most likely to be finally chosen (set to favor privacy over accuracy), so that we could isolate the distortions and demonstrate the effectiveness of various methodological modifications. One cannot draw conclusions about the accuracy of the data the Census Bureau will release for the 2020 Census based on these demonstration products.
76. Dr. Barber is premature in drawing conclusions about the accuracy of the 2020 redistricting data before the Census Bureau has set a final privacy-loss budget, and he is further incorrect in opining on the accuracy of differential privacy without considering the relative error of alternatives. Dr. Barber focuses most of his report on the possible quality concerns of differentially private 2020 Census data releases with no attention to (1) the demonstrated privacy risks of a 2020 Census protected by legacy methods and (2) the accuracy of alternatives to differential privacy including enhanced swapping or suppression. As I show in this declaration, all disclosure avoidance systems trade-off accuracy for confidentiality protection. They must be compared to each other. Releasing the redistricting data without disclosure avoidance procedures – tabulating the Census Edited File directly – is not an option and was not done for the 1990, 2000, or 2010 Censuses.
77. Dr. Barber relies on external studies that draw incorrect conclusions and use early demonstration data products. In his declaration, Dr. Barber quotes Santos-Lozada, et al. (2020) on page 14 by saying that “[i]nfusing noise in the data, in comparison to the

current disclosure avoidance system, will produce inaccurate patterns of demographic change with higher levels of error found in the calculations for non-Hispanic blacks and Hispanics. At the same time, these counts are bound to impact post-2020 districting for both federal and state elections, as well as evaluations of that redistricting. . . . [T]hese changes in population counts will affect understandings of health disparities in the nation, leading to overestimates of population-level health metrics of minority populations in smaller areas and underestimates of mortality levels in more populated ones.” The Santos-Lozada et al. paper uses the October 2019 Demonstration Data Product. Therefore, its conclusions are only applicable to the state of the algorithms and the overall privacy-loss budget used for that early release. Those were neither the final algorithms nor the final privacy-loss budget. I informed the editors of the Proceedings of the National Academy of Sciences of these defects during the peer-review process. I strongly recommended that the word “will” in the title be changed to “may” for these reasons. There is nothing statistically incorrect in the paper except for the general failure of these demographers to account for estimation error due to disclosure avoidance when doing their statistical analyses as I have noted in my own scholarly work⁵⁹ and other statisticians and computer scientists have also noted.⁶⁰ The fatal error in the Santos-Lozada et al. paper is drawing conclusions from preliminary data generated by an obsolete version of the 2020 Census DAS using obsolete settings for the privacy-loss budget and its allocation. Those conclusions are wrong and so, by extension, are those of Dr. Barber.

⁵⁹ Abowd, John M. and Ian Schmutte “Economic Analysis and Statistical Disclosure Limitation” *Brookings Panel on Economic Activity* (Spring 2015): 221-267. [[download article and discussion](#), open access] [[download preprint](#)].

⁶⁰ Wasserman L. and S. Zhou “A Statistical Framework for Differential Privacy,” *Journal of the American Statistical Association*, Vol. 105, No. 489 (2010):375-389, DOI: [10.1198/jasa.2009.tm08651](https://doi.org/10.1198/jasa.2009.tm08651).

78. Dr. Barber's conclusions do not take into account that if the Census Bureau were forced to hold the number of people in housing units invariant at the block level, that would, in turn, require adding more noise and error to the demographic characteristics of those individuals in an effort to offset what amounts to assigning block-level populations an infinite privacy-loss budget. As I show in my declaration, doing so is unnecessary and harmful to both accuracy and confidentiality protection. The correct procedure is to set accuracy targets for meaningful aggregations then tune the disclosure avoidance procedures to meet them. This procedure is transparent when using differential privacy, but it was also done for the 2010 swapping system albeit in memos that are also protected by 13 U.S. Code §§ 8(b) & 9.
79. Furthermore, Dr. Barber's work draws incorrect conclusions about biases in rural areas and for specific small populations. In his declaration, Dr. Barber states on page 13 that "[p]laces with fewer people (rural locations) and areas with smaller, distinctive populations (minority communities) are more likely to be impacted since these are the places where identification is more concerning, and the application of statistical noise is more likely to have a larger impact on the summary statistics derived from the altered data." He concludes on pages 13 and 14 that "...the process of differential privacy is not applied equally across the entire population. Places with fewer people (rural locations) and areas with smaller, distinctive populations (minority communities) are more likely to be impacted since these are the places where identification is more concerning, and the application of statistical noise is more likely to have a larger impact on the summary statistics derived from the altered data." This conclusion is incorrect. His analysis should say that the privacy-loss of the respondents in these small areas is being treated equally and identically to the privacy-loss of the respondents in large population areas; that is, every single respondent gets the full privacy protection afforded by the DAS—unlike the 2010 system, which only tried to protect certain households. To properly compare urban/rural statistics before and after the

application of disclosure avoidance, regardless of the system, the full algorithm assigning rural/urban status must be used on both the privacy-protected and confidential data. Dr. Barber has not done this.

80. Dr. Barber's work makes incorrect assertions about the non-negativity constraint. In his declaration, Dr. Barber cites Riper, Kugler, and Ruggles (2020) on page 13 stating that "[t]he non-negativity constraint requires that every cell in the final detailed histogram be non-negative. As described above, many of the cells in the noisy household histograms will be negative, especially for geographic units with smaller numbers of households. Returning these cells to zero effectively adds households to these small places, resulting in positive bias." This point is not an accurate description of how non-negativity is being handled in the post-processing of the noisy histogram. The analysis should say that negative values are not simply being returned to zero, but that all blocks with housing units are used to estimate the population counts subject to a non-negativity constraint on the solutions. That is, negative values are not "[r]eturning to zero," the entire 2,016 element matrix (for the redistricting data) is smoothed to a consistent, non-negative matrix for each of the 8,000,000 blocks, 275,000 block groups, 75,000 tracts, 3,143 counties, 51 states (including DC), and the U.S. simultaneously.⁶¹ At the block-level, there are expected to be an average of only 40 people represented across the 2,016 cells. This is the inherent sparsity that any disclosure avoidance system must address. Dr. Barber claims on page 13 that "[t]he combination of the non-negativity constraint and population invariants consistently leads to bias increasing counts of small subgroups and small geographic units and decreasing counts of larger subgroups and geographic units." While the statement is correct in

⁶¹ The matrix is 2,016 elements rather than 252 because there are eight elements in the Group Quarters Table P5 (seven group quarter types and "not a group quarters") that also interact with the other categories. The number of geographic entities at each level is based on approximate values for 2020 tabulation geographies.

principle, the magnitudes shown in his report are not representative of the final re-districting data product. At the levels of privacy-loss budget used for the forthcoming April 2021 Demonstration Data Product, the consequences of the non-negativity constraint were tightly controlled for population areas of at least 500 total persons. The remaining variation in block-level statistics, including small biases, is required to protect locational privacy and deliver consistent data. It is well within the inherent variability of block-level census data, as shown in my declaration.

81. Dr. Barber argues that the amount of error observed in the demonstration files indicates that differential privacy cannot produce data sufficient for important use cases. Mr. Barber's focus on the percentage of blocks in the demonstration data that differ at all from the official 2010 Census data (even if that difference represents the addition or subtraction of a single individual from the block) ignores two important points. First, the entire objective of our implementation of differential privacy is to infuse sufficient noise in block-level data to protect against reconstruction-abetted re-identification attacks while ensuring that when those blocks are aggregated into larger geographies of interest (voting districts, towns, etc.) those relative errors diminish and the accuracy of the tabulations improves. Second, the overall accuracy of the data is a direct consequence of the global privacy-loss budget selected and how it is allocated. The demonstration data used by both Dr. Barber and Mr. Bryan for their analyses, which use a substantially lower privacy-loss budgets than will be used for the final 2020 Census data products, can therefore be expected to be substantially "noisier" than the final data will be. Examples of noise levels in the April 2021 Demonstration Data Product provided in my report and verifiable when those data are released later this month confirm my claims.

82. Mr. Bryan assesses the accuracy of the four Demonstration Data Products (October 2019, May 2020, September 2020 and November 2020) using the percent of blocks with any change at all (pp. 9-13) or percentage errors (pp. 16-19). Both sets of analyses are

based on obsolete versions of the DAS, but they also make serious errors that will still be salient when he uses the April 2021 Demonstration Data Product. The DAS was designed to control the error in counts, not percentages. The basic tables in the P.L. 94-171 Redistricting Data Summary File are counts of resident persons living in specific geographies who have features chosen from the following taxonomy {any age, voting age}, {Hispanic/Latino, not Hispanic/Latino}, and any combination of {Afro-American/Black, American Indian/Alaska Native, Asian, Native Hawai'ian/Pacific Islander, White, Some other race} except "none." The specific aggregate geographies available in the data product are all built from census blocks, but it is the counts of persons in those aggregate geographies, including voting districts, not the block counts themselves that must be accurate enough to be fit for redistricting. Block-level errors, whether in counts or percentages, are irrelevant except to the extent that they are not controlled in larger-population geographies. In 2010, the average population in a block was 28 and the average population in an occupied block was 49. Any block-level variation in one of the 2,016 cells of the redistricting data for total populations this small is going to appear as a "large" percentage error. Indeed, most of those statistics have a base of zero, making percentage variation undefined and meaningless. The DAS must introduce noise into the block-level data to achieve any confidentiality protection at all. This statement is also true for the systems that were used in the 1970 to 2010 Census. The noise from suppression (1970, 1980) is counts that are simply not reported at the block level. The noise from blank and impute (1990) is due to the imputation modeling. The noise from swapping (2000, 2010) is due the exchange of geographic identifiers across blocks. All confidentiality protection applied to block-level redistricting data produces errors of the sort described by Mr. Bryan. Furthermore, many of the supposed DAS errors in Mr. Bryan's analysis cancel out when blocks are aggregated into larger-population geographies like block groups, census tracts, towns, counties, and congressional districts. This is not an accident; it is a carefully

designed feature of the DAS. The tabulation of the protected microdata might miss a person in one block, but have an “excess” person in the neighboring block for a particular characteristic. Because the DAS uses direct measurements from the U.S. all the way down to the block to estimate the counts at every level of geography, whether on- or off-spine, they are all much more accurate than any of the block estimates that comprise them. This is easy to see in any balanced summary of the accuracy of the DAS. Counties and places have far smaller percentage errors than the average percentage error of the blocks that compose them.

CLARIFYING STATEMENT QUOTED IN COMPLAINT

83. Plaintiffs assert, quoting an article in 2018 by the demographer Steven Ruggles and others, that I claimed that database reconstruction does not pose a significant re-identification threat. I made the statement that plaintiffs reference indirectly at the December 14, 2018 meeting of the Federal Economic Statistics Advisory Committee (FESAC) in my own presentation.⁶² Dr. Ruggles was on the FESAC program in the same session. I made the remarks in December 2018 as a report on ongoing research.⁶³ At the February 16, 2019 session of the American Association for the Advancement of Science (AAAS), I retracted my tentative conclusion about re-identification based on additional research reported there. The full text and presentation of the AAAS session are attached as Appendices H and I.⁶⁴ To be clear, the Census Bureau’s simulated recon-

⁶² Federal Economic Statistics Advisory Committee program: [FESAC Meeting Agenda December 2018 \(bea.gov\)](https://www.bea.gov/fesac/2018/12/14/fesac-meeting-agenda-december-2018).

⁶³ My remarks at the December 18, 2018 FESAC: [Microsoft PowerPoint - Abowd Presentation \(bea.gov\)](https://www.bea.gov/fesac/2018/12/18/microsoft-powerpoint-abowd-presentation).

⁶⁴ AAAS materials for the February 16, 2019 session area also here: <https://blogs.cornell.edu/abowd/files/2019/04/2019-02-16-Abowd-AAAS-Talk-Saturday-330-500->

struction attack on the 2010 Census data described in this declaration and in the accompanying appendix materials shows there is a significant re-identification risk. However, the Census Bureau's Data Stewardship Executive Policy Committee (DSEP) acted to adopt differential privacy as soon as that research showed that an accurate microdata reconstruction was feasible. It did not require, nor should it have required, the subsequent demonstration that those reconstructed microdata permit between 52 and 179 million correct re-identifications from the 2010 Census. The reconstructed microdata fail the *2010 Census* microdata disclosure avoidance requirements—the requirements that were in place for that census—because they contain geographic identifiers (the block code) that relate to a minimum population of one rather than the 100,000 person minimum population that contemporary standards required. The reconstructed microdata also did not impose any of the minimum population thresholds required of the tabulation variables, especially age.⁶⁵ These requirements were already in place because it is well understood at the Census Bureau and in the official statistics community worldwide that geographic identifiers for low-population areas, sex, and exact age in microdata files are a major disclosure risk especially in population censuses.

IMPACT OF ANY COURT RULING BARRING USE OF DIFFERENTIAL PRIVACY

84. Were the Court to rule that the Census Bureau was precluded from using differential privacy for the 2020 Census P.L. 94-171 Redistricting Data Summary File, we would be faced with hard choices. The inevitable result would be significant delay in deliv-

[session-FINAL-as-delivered-2jr4lzb.pdf](#) and <https://blogs.cornell.edu/abowd/files/2019/04/2019-02-16-Abowd-AAAS-Slides-Saturday-330-500-session-FINAL-as-delivered-1iqsdg2.pdf>.

⁶⁵ McKenna (2019a).

ery of the already-delayed redistricting data and diminished accuracy. Either the Census Bureau would have to revert to using suppression (as was last used in the 1980 Census) or use enhanced swapping (as was used in the 1990 to 2010 Censuses, but at a much higher rate and with fewer invariants). Either choice would delay results and diminish accuracy.

85. The effect on the schedule for delivering redistricting data would be substantial. The Census Bureau cannot ascertain the length of the delay until it understands any parameters the Court might place on its choice of methodology, but under all scenarios the delay would be multiple months. This delay is unavoidable because the Census Bureau would need to develop and test new systems and software, then use them in production and subject the results to expert subject matter review prior to production of data. The Census Bureau has been developing the systems and software to use differential privacy for several years—the agency has spent millions of dollars purchasing cloud computer capacity and writing and tuning code. The systems and software are ready to go and await only final tuning and a decision on the privacy-loss budget.

86. Even if the agency was ordered to repeat exactly what was done in 2010 (despite the serious risks to privacy the Census has identified), we could not simply “flip a switch” and revert to the prior methodology. Instead, we would need to conduct the requisite software development and testing. The 2020 Census’s system architecture is completely different than that used in the 2010 Census, and it is thus not possible to simply “plug in” the disclosure-avoidance system used in 2010.

87. Not only would redistricting data be further delayed, but the resulting data would be less accurate. Both swapping and suppression are blunt instruments for privacy protection. Unlike differential privacy, neither can be effectively tuned to optimize for data accuracy. Knowing that the 2010 Census results were vulnerable to reconstruction, the Census Bureau cannot simply repeat the swapping protocols from the 2010

census, but rather would be forced to fashion appropriate levels of protection for either system. Using an appropriate level of protection for either suppression or swapping would produce far less accurate data than would differential privacy.

88. I would urge any court to be quite wary of opining on the suitability of particular methods for conducting disclosure avoidance, as these decisions are highly technical and can have unanticipated consequences. The only reason the Court knows so much about the proposed methods for the 2020 Census is that transparency does not undermine their confidentiality protections, which is not the case for either swapping or suppression. While we cannot predict the full impact of any change, there is a danger than any change would have cascading effects on data accuracy and privacy, making race and ethnicity data, along with age data, substantially less accurate. Any sort of change in the basic methodology would be minimally tested and would not have the benefit of any input from the user community.

89. In conclusion, it is my professional opinion that the Census Bureau's Data Stewardship Executive Policy Committee should be permitted to control the type and parameters of any disclosure avoidance system used for the 2020 Census, just as it did for the 2010 Census and just as its predecessor committees did for decennial censuses conducted since the passage of the Census Act (13 U.S. Code) in 1954.

I declare under penalty of perjury that the foregoing is true and correct.

DATED and SIGNED:

JOHN ABOWD

Digitally signed by JOHN ABOWD
Date: 2021.04.13 08:45:14 -04'00'

John M. Abowd

Chief Scientist and Associate Director for Research and Methodology

United States Bureau of the Census

John M. Abowd

Home Professional Information Courses Recent News Special Materials

Professional Information

[Updated April 1, 2021]

CONTACT INFORMATION

U.S. Census Bureau

HQ 8H120 ATTN: Sara Sullivan

4600 Silver Hill Road

Private delivery services (FedEx, UPS, etc.) physical location: Suitland, MD 20746

USPS mail only: Washington, DC 20233

Voice: +1.301.763.5880

Mobile: +1.202.591.0766

Fax: +1.301.763.8360

Executive assistant Sara Sullivan: +1.301.763.5116

E-mail: john.maron.abowd@census.gov

ILR School

USPS mail only (send private delivery service items to the address above):

275 Ives Hall

Cornell University

Ithaca, New York 14853-3901

Assistant: LDI@cornell.edu

E-mail: john.abowd@cornell.edu

Webpage: <https://blogs.cornell.edu/abowd/> or <https://www.johnabowd.com>

Twitter: [@john_abowd](https://twitter.com/john_abowd) (opinions are my own)

Short biography in PDF format

CURRENT POSITIONS

Chief Scientist and Associate Director for Research and Methodology, U. S. Census Bureau, IPA June 1, 2016 – March 27, 2020; Career Senior Executive Service March 29, 2020 –

Edmund Ezra Day Professor, Department of Economics, Cornell University, July 2011 – currently on leave

Director, Labor Dynamics Institute, Cornell University, October 2011 – currently on leave

Founding member and Professor of Information Science (by courtesy), Faculty of Computing and Information Science, July 2000 – currently on leave

Professor of Statistics and Data Science, September 2013 – currently on leave

Member of the Graduate Fields of Economics, Industrial and Labor Relations, Information Science, and Statistics

Search ...

SEARCH

INSTITUTIONS

U.S. Census Bureau

Cornell Economics

Labor Dynamics Institute

NCRN node at Cornell

CISER

OTHER INFORMATION

Google Scholar

ORCID

RePEC/Ideas

SSRN

Research Associate, [National Bureau of Economic Research](#), 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138, September 1983 – (on leave while serving at the U.S. Census Bureau)

Research Affiliate, [Centre de Recherche en Economie et Statistique/INSEE](#), 15, bd Gabriel Péri, 92245 Malakoff Cedex France, November 1997 –

Research Fellow, [IZA \(Institute for the Study of Labor\)](#), P.O. Box 7240 D-53072 Bonn, Germany, June 2002 –

Research Fellow, [IAB \(Institut für Arbeitsmarkt-und Berufsforschung\)](#), Dienstgebäude Weddigenstraße 20-22, 90478 Nürnberg, Germany, January 2013 –

President and Principal, ACES-Research, LLC, john@aces-research.com, July 2007 –

Editor, [Journal of Privacy and Confidentiality](#) Online journal

PREVIOUS AND VISITING POSITIONS

Distinguished Senior Research Fellow, [United States Census Bureau](#), September 1998 – May 2016

[Associate Chair](#), [Department of Economics](#), Cornell University, August 2015 – May 2016

Visiting Professor, [Center for Labor Economics](#), University of California-Berkeley, August 2014 – July 2015

[Director of Graduate Studies, Economics](#), July 2010 – June 2014

Professor of Economics and Econometrics, University of Notre Dame, January 2008 – May 2008.

Director, [Cornell Institute for Social and Economic Research \(CISER\)](#), July 1999 – December 2007

Associate Director, Cornell Theory Center (became [Cornell University Center for Advanced Computing](#)), September, 2006 – August 2007.

Professor of Labor Economics, Cornell University, January 1990 – October 2001.

Edmund Ezra Day Professor, School of Industrial and Labor Relations, November 2001 –

Associate Director, Cornell Institute for Social and Economic Research (CISER), July 1998 – June 1999.

Chair, Department of Labor Economics, Cornell University, September 1992 – June 1998.

Acting Director, CISER, January 1998-June 1998.

Professeur invité, Laboratoire de Microéconomie Appliquée-Theorie Et Applications en Microéconomie et macroéconomie (LAMIA-TEAM), Université de Paris-I (Panthéon-Sorbonne), May 1998.

Consultant, Centre de Recherche en Economie et Statistique (CREST), Institut National de la Statistique et des Etudes Economiques (INSEE), February 1997.

Professeur invité, ERMES (Equipe de Recherche sur les Marchés, l'Emploi et la Simulation) Université Panthéon-Assas (Paris II), October 1995 – July 1996 (part time).

Professor, Samuel Curtis Johnson Graduate School of Management, Cornell University (adjunct appointment), August 1987 – July 1995.

Chercheur étranger, Institut National de la Statistique et des Etudes Economiques (INSEE), Paris, Department of Research, August 1991 – July 1992, January 1993, January 1994.

Professeur visitant, HEC (Hautes Etudes Commerciales, Paris) Department of Finance and Economics, September 1991 – July 1992 and January 1993, December 1993 – January 1994.

Professeur visitant, CREST (Centre de Recherche en Statistiqu et Economie, Paris), September 1991 – July 1992, July 1993.

Associate Professor with tenure, Cornell University, August 1987 – December 1989.

Research Associate, Industrial Relations Section, Department of Economics, Princeton University, September 1986 – August 1987.

Visiting Associate Professor of Economics, Department of Economics, Massachusetts Institute of Technology, September 1985 – August 1986.

Associate Professor of Econometrics and Industrial Relations, Graduate School of Business, University of Chicago, September 1982 – August 1986. Assistant Professor, September 1979 – August 1982. Visiting Assistant Professor, September 1978 – August 1979.

Senior Study Director/Research Associate, NORC/Economics Research Center, 6030 Ellis Avenue, Chicago, Illinois 60637, September 1978 – August 1986.

Academic Consultant, Centre for Labour Economics, London School of Economics, January 1979 – April 1979.

Assistant Professor of Economics, Department of Economics, Princeton University, September 1977 – August 1979 (on leave September 1978 – August 1979). Lecturer in Economics, September 1976 – August 1977.

Associate Editor, *Journal of Business and Economic Statistics*, 1983 – 1989.

Editorial Board, *Journal of Applied Econometrics*, 1987 – 1989.

Associate Editor, *Journal of Econometrics*, 1987 – 1989.

EDUCATION

Ph.D. Department of Economics, University of Chicago, December 1977.
Thesis: An Econometric Model of the U.S. Market for Higher Education

M.A. Department of Economics, University of Chicago, March 1976.

A.B. Department of Economics (with highest honors), University of Notre Dame, May 1973.

LANGUAGES

English (native), French

HONORS AND FELLOWSHIPS

Fellow, American Association for the Advancement of Science (elected October 2020)

Julius Shiskin Award, American Statistical Association, Business and Economic Statistics Section (2016)

Cornell University, **Graduate and Professional Student Assembly Award for Excellence in Teaching, Advising, and Mentoring** (May 2015)

Fellow, *Econometric Society* (elected November 2014)

Roger Herriot Award, American Statistical Association, Government and Social Statistics Sections (2014)

Elected member, *International Statistical Institute* (March 2012)

Council of Sections (2014-2016), Chair (2013) *Business and Economic Statistics Section* (Chair-elect 2012), American Statistical Association

President (2014-2015), *Society of Labor Economists*, President-elect (2013-2014), Vice President (2011-2013)

Fellow, *The American Statistical Association* (elected August 2009)

Fellow, *Society of Labor Economists* (elected November 2006)

La bourse de haut niveau du Ministère de la Recherche et de la Technologie, fellowship for research at the Institut National de la Statistique et des Etudes Economiques (INSEE) awarded by the French Government, September 1991 – February 1992.

National Institute of mental Health postdoctoral fellow at NORC, September 1978 – August 1980.

National Institute of Mental Health pre-doctoral fellow at the University of Chicago, September 1973 – June 1976.

TEACHING EXPERIENCE

Graduate:

Microeconometrics using Linked Employer-Employee Data (CREST-ENSAE)
Understanding Social and Economic Data (Cornell, co-instructor: Lars Vilhuber)
 Third-year Research Seminar I and II (Cornell)
 Seminar in Labor Economics I, II, and III (Cornell)
Microéconometrie des Données Appariées (CREST-GENES, in French)
Microéconomie et Microéconometrie du Travail (Université de Paris I, in French)
Economie du Travail (Université de Paris II, in French)
Economics of Compensation and Organization (Cornell)
International Human Resource Management (Cornell)
Corporate Finance (Hautes Etudes Commerciales, Paris)
International Human Resource Management (HEC, Paris)
Workshop in Labor Economics (Cornell)
Economics of Collective Bargaining (Cornell)
Executive Compensation (Cornell)
Labor Economics (MIT)
Labor and Public Policy (MIT)
Applied Econometrics I, II (Chicago)
Introduction to Industrial Relations (Chicago)
Econometric Theory I (Chicago)
Industrial Relations and International Business (Chicago)
Workshop in Economics and Econometrics (Chicago)
Econometric Analysis of Time Series (Princeton)
Mathematics for Economists (Princeton)

Undergraduate:

Understanding Social and Economic Data (Cornell, co-instructor: Lars Vilhuber)
Introductory Microeconomics (Cornell)
Economics of Employee Benefits (Cornell)
Economics of Wages and Employment (Cornell)

Corporate Finance (Cornell)
 Introduction to Econometrics (Princeton)
 Microeconomics (Princeton)

BIBLIOGRAPHY

Books

1. Abowd, John M. and Francis Kramarz (eds.) *The Microeconometrics of Human Resource Management*, special issue of *Annales d'économie et de statistique* 41/42 (Paris: ADRES, January/June 1996).
2. Abowd, John M. and Richard B. Freeman (eds.) *Immigration, Trade and the Labor Market* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 1991).

Articles

1. McKinney, Kevin L., John M. Abowd, and John Sabelhaus, "United States Earnings Dynamics: Inequality, Mobility, and Volatility," In Raj Chetty, John N. Friedman, Janet C. Gornick, Barry Johnson, and Arthur Kennickel, eds., *Measuring the Distribution and Mobility of Income and Wealth*, (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2021), forthcoming. [[download preprint](#)] [[download chapter \(open access\)](#)]
2. Abowd, John M. "Official Statistics at the Crossroads: Data Quality and Access in an Era of Heightened Privacy Risk," *The Survey Statistician*, Vol. 83 (January 2021):23-26. [[download \(open access\)](#)]
3. McKinney, Kevin L., Andrew S. Green, Lars Vilhuber and John M. Abowd "Total Error and Variability Measures for the Quarterly Workforce Indicators and LEHD Origin-Destination Employment Statistics in OnTheMap" *Journal of Survey Statistics and Methodology* (November 2020). [[download arxiv preprint](#)], DOI: <https://doi.org/10.1093/jssam/smaa029>, supplemental online materials DOI: <https://doi.org/10.5281/zenodo.3951670>
4. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber "Why the Economics Profession Must Actively Participate in the Privacy Protection Debate," *American Economic Association: Papers and Proceedings*, Vol. 109 (May 2019): 397-402, DOI:10.1257/pandp.20191106. [[download preprint](#)]
5. Abowd, John M. and Ian M. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review*, Vol. 109, No. 1 (January 2019):171-202, DOI:10.1257/aer.20170627. [AER, [ArXiv preprint](#), [Replication information](#)]
6. Weinberg, Daniel H., John M. Abowd, Robert F. Belli, Noel Cressie, David C. Folch, Scott H. Holan, Margaret C. Levenstein, Kristen M. Olson, Jerome P. Reiter, Matthew D. Shapiro, Jolene Smyth, Leen-Kiat Soh, Bruce D. Spencer, Seth E. Spielman, Lars Vilhuber, and Christopher K. Wikle "Effects of a Government-Academic Partnership: Has the NSF-Census Bureau Research Network Helped Secure the Future of the Federal Statistical System?" *Journal of Survey Statistics and Methodology* (2018) DOI:10.1093/jssam/smy023. [[download](#), [open access](#)] [[download preprint](#)]
7. Abowd, John M., Ian M. Schmutte and Lars Vilhuber "Disclosure Limitation and Confidentiality Protection in Linked Data," in A.Y. Chun, M. Larson, J. Reiter, and G. Durrant (eds.) *Administrative Records for Survey Methodology* (New York: Wiley, forthcoming). [[download preprint](#)]
8. Abowd, John M., Kevin L. McKinney and Ian M. Schmutte "Modeling Endogenous Mobility in Earnings Determination," *Journal of Business and Economic Statistics* Vol. 37, Issue 3 (2019):405-418. DOI: [10.1080/07350015.2017.1356727](https://doi.org/10.1080/07350015.2017.1356727). [[download preprint](#)] [JBES]
9. Abowd, John M., Francis Kramarz, Sébastien Perez-Duarte, and Ian Schmutte "Sorting between and within Industries: A Testable Model of Assortative Matching," *Annals of Economics and Statistics* 129 (March 2018): 1-32. NBER WP-20472. [[download preprint](#)] [[programs](#)] [[data](#)]
10. Abowd, John M., Kevin L. McKinney and Nellie Zhao "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data," *Journal of Labor Economics* 36, S1 (January 2018):S183-S300 DOI: 10.1086/694104. [[download](#), [not copyrighted](#)] [[download preprint](#)]

11. Abowd, John M. "How Will Statistical Agencies Operate When All Data Are Private?" *Journal of Privacy and Confidentiality*, Vol. 7, Issue 3, Article 1 (2017). [[download](#), [open journal](#)]
12. Haney, Samuel, Ashwin Machanavajjhala, John M. Abowd, Matthew Graham, Mark Kutzbach, and Lars Vilhuber "Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics," *ACM SIGMOD 2017*, DOI: 10.1145/3035918.3035940. [[download](#)]
13. Abowd, John M. and Kevin L. McKinney "Noise Infusion as a Confidentiality Protection Measure for Graph-based Statistics" *Statistical Journal of the International Association for Official Statistics* (2016) Vol. 32, No. 1, pp. 127-135, DOI: 10.3233/SJI-160958. [[download article](#), [open access](#)] [[download preprint](#)]
14. Abowd, John M. and Ian Schmutte "Economic Analysis and Statistical Disclosure Limitation" *Brookings Panel on Economic Activity* (Spring 2015): 221-267. [[download article and discussion](#), [open access](#)] [[download preprint](#)]
15. Schneider, Matthew J. and John M. Abowd "A New Method for Protecting Interrelated Time Series with Bayesian Prior Distributions and Synthetic Data," *Journal of the Royal Statistical Society, Series A* (2015) DOI:10.1111/rssa.12100. [[download preprint](#)]
16. Lagoze, Carl, William C. Block, Jeremy Williams, Lars Vilhuber, and John M. Abowd "Data Management of Confidential Data." *International Journal of Digital Curation* 8, no. 1 (2013): 265-278. doi:10.2218/ijdc.v8i1.259. [[download preprint](#)]
17. Abowd, John M. and Martha H. Stinson "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *Review of Economics and Statistics*, Vol. 95, No. 5 (December 2013): 1451-1467. doi:10.1162/REST_a_00352. [[download](#), [not copyrighted](#)]
18. Abowd, John M., Matthew J. Schneider and Lars Vilhuber "Differential Privacy Applications to Bayesian and Linear Mixed Model Estimation," *Journal of Privacy and Confidentiality*. Vol. 5: Iss. 1 (2013): Article 4. [[download](#), [open access](#)]
19. Abowd, John M., Francis Kramarz, Paul Lengermann, Kevin L. McKinney, and Sébastien Roux "Persistent Inter-Industry Wage Differences: Rent Sharing and Opportunity Costs," *IZA Journal of Labor Economics*, 2012, 1:7, doi:10.1186/2193-8997-1-7. [[download](#), [open access](#)] [[online Appendix](#)]
20. Abowd, John M., Lars Vilhuber and William Block "A Proposed Solution to the Archiving and Curation of Confidential Scientific Inputs," in J. Domingo-Ferrer and I. Tinnirello, eds., *Privacy in Statistical Databases 2012*, LNCS 7556, pp. 216-225, (2012). [[download](#), [open access](#)]
21. Abowd, John M. and Lars Vilhuber "Did the Housing Price Bubble Clobber Local Labor Markets When It Burst?" *American Economic Review Papers and Proceedings* Vol. 102, No. 3 (May 2012): 589-93, doi:pdfplus/10.1257/aer.102.3.589. [[download preprint](#)] [[online Appendix](#)] [[data Readme](#)] [[data](#)]
22. Abowd, John M., R. Kaj Gittings, Kevin L. McKinney, Bryce E. Stephens, Lars Vilhuber, and Simon Woodcock "Dynamically Consistent Noise Infusion and Partially Synthetic Data As Confidentiality Protection Measures for Related Time-series," Federal Committee on Statistical Methodology, Office of Management and Budget, 2012 Research Conference Papers. [[download](#), [open access](#), cited on May 21, 2012] [[download archival copy](#)].
23. Abowd, John M. and Matthew Schneider "An Application of Differentially Private Linear Mixed Modeling," *ICDMW*, pp. 614-619, 2011 IEEE 11th International Conference on Data Mining Workshops, 2011. [[download](#), [open access](#)]
24. Kinney, Satkartar K., Jerome P. Reiter, Arnold P. Reznick, Javier Miranda, Ron S. Jarmin, and John M. Abowd "Towards Unrestricted Public Use Business Micro-data: The Synthetic Longitudinal Business Database," *International Statistical Review*, Vol. 79, No. 2 (December 2011): 362-84, doi:10.1111/j.1751-5823.2011.00153.x. [[download](#), [subscription required](#)] [[download preprint](#)]
25. Abowd, John M. and Lars Vilhuber "National Estimates of Gross Employment and Job Flows from the Quarterly Workforce Indicators with Demographic and Industry Detail," *Journal of Econometrics*, Vol. 161 (March 2011): 82-99, doi: 10.1016/j.jeconom.2010.09.008. [[download preprint](#)] [[data](#)]
26. Abowd, John M., Bryce Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators" in T. Dunne, J.B. Jensen and M.J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2009), pp. 149-230. [[download](#), [not copyrighted](#)] [[archival copy](#)]

27. Abowd, John M., Kevin McKinney and Lars Vilhuber "The Link between Human Capital, Mass Layoffs, and Firm Deaths" in T. Dunne, J.B. Jensen and M.J. Roberts, eds., *Producer Dynamics: New Evidence from Micro Data* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2009), pp. 447-472. [[download, not copyrighted](#)] [[archival copy](#)]
28. Abowd, John M. and Lars Vilhuber "How Protective are Synthetic Data," in J. Domingo-Ferrer and Y. Saygun, eds., *Privacy in Statistical Databases*, (Berlin: Springer-Verlag, 2008), pp. 239-246. [[download preprint](#)]
29. Abowd, John M., Francis Kramarz and Simon Woodcock "Econometric Analyses of Linked Employer-Employee Data," in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data* (The Netherlands: Springer, 2008), pp. 727-760. [[download preprint](#)]
30. Abowd, John M., John Haltiwanger and Julia Lane "Wage Structure and Labor Mobility in the United States," in E. P. Lazear and K. L. Shaw, eds., *Wage Structure, Raises, and Mobility: International Comparisons of the Structure of Wages within and Across Firms* (Chicago: University of Chicago Press for the National Bureau of Economic Research, 2008), pp. 81-100. [[download](#)] [[download preprint](#)]
31. Machanavajjhala Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber "Privacy: Theory Meets Practice on the Map," International Conference on Data Engineering (ICDE) 2008: 277-286, doi:10.1109/ICDE.2008.4497436. [[download preprint](#)]
32. Abowd, John M. and Francis Kramarz "Human Capital and Worker Productivity: Direct Evidence from Linked Employer-Employee Data," *Annales d'Economie et de Statistique*, No. 79/80, (Juillet/Décembre 2005): 323-338. [[download preprint](#)]
33. Torra, V. J.M. Abowd and J. Domingo-Ferrer "Using Mahalanobis Distance-Based Record Linkage for Disclosure Risk Assessment," in J. Domingo-Ferrer and Luisa Franconi (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2006), pp. 233-242. [[download preprint](#)]
34. Abowd, John M., Francis Kramarz and Sébastien Roux "Wages, Mobility and Firm Performance: Advantages and Insights from Using Matched Worker-Firm Data," *Economic Journal*, Vol. 116, (June 2006): F245-F285. [[download preprint](#)]
35. Abowd, John M., Francis Kramarz and Sébastien Roux "Heterogeneity in Firms' Wages and Mobility Policies," in H. Bunzel, B.J. Christensen, G.R. Neumann and J-M. Robin, eds., *Structural Models of Wage and Employment Dynamics*, (Amsterdam: Elsevier Science, 2006), pp. 237-268. [[download preprint](#)]
36. Abowd, John M. and Lars Vilhuber "The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers," *Journal of Business and Economics Statistics*, Vol. 23, No. 2 (April 2005): 133-152, *JBES* Joint Statistical Meetingsinvited paper with discussion and "Rejoinder" (April 2005): 162-165. [[download preprint](#)].
37. Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky "The Relation among Human Capital, Productivity and Market Value: Building Up from Micro Evidence," in *Measuring Capital in the New Economy*, C. Corrado, J. Haltiwanger, and D. Sichel (eds.), (Chicago: University of Chicago Press for the NBER, 2005), Chapter 5, pp. 153-198. [[download, not copyrighted](#)] [[download preprint](#)]
38. Abowd, John M. and Simon Woodcock "Multiply-Imputing Confidential Characteristics and File Links in Longitudinal Linked Data," in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2004), pp. 290-297. [[download preprint](#)]
39. Abowd, John M., John Haltiwanger and Julia Lane "Integrated Longitudinal Employee-Employer Data for the United States," *American Economic Review Papers and Proceedings*, Vol. 94, No. 2 (May 2004): 224-229. [[download preprint](#)]
40. Abowd, John M. and Julia Lane "New Approaches to Confidentiality Protection: Synthetic Data, Remote Access and Research Data Centers," in J. Domingo-Ferrer and V. Torra (eds.) *Privacy in Statistical Databases* (Berlin: Springer-Verlag, 2004), pp. 282-289. [[download preprint](#)]
41. Abowd, John M. and Francis Kramarz "The Costs of Hiring and Separations," *Labour Economics*, Vol. 10, Issue 5 (October 2003): 499-530. [[download preprint](#)]
42. Abowd, John M. "Unlocking the Information in Integrated Social Data," *New Zealand Economic Papers*, 0077-9954, Vol. 36, No. 1 (June 2002): 9-31. [[download preprint](#)]
43. Abowd, John M. and Orley Ashenfelter "Using Price Indices and Sale Rates to Assess Short Run Changes in the Market for Impressionist and Contemporary Paintings" in *The Economics of Art Auctions*, G. Mosetto and M. Vecco (eds.), (Milan: F. Angeli Press, 2002). [[download preprint](#)] [[access book](#)]

44. Abowd, John M. and Simon Woodcock "Disclosure Limitation in Longitudinal Linked Data," in *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, P. Doyle, J. Lane, J. Theeuwes, and L. Zayatz (eds.), (Amsterdam: North Holland, 2001), 215-277. [[download preprint](#)]
45. Abowd, John M., Bruno Crépon and Francis Kramarz "Moment Estimation with Attrition: An Application to Economic Models," *Journal of the American Statistical Association*, 96, No. 456 (December 2001): 1223-1231. [[download preprint](#)]
46. Abowd, John M., Francis Kramarz, David Margolis, and Kenneth Troske "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," *Journal of the Japanese and International Economies*. Vol. 15, No. 4, (December 2001): 419-436. [[download preprint](#)]
47. Abowd, John M. Julia Lane and Ronald Prevost "Design and Conceptual Issues in Realizing Analytical Enhancements through Data Linkages of Employer and Employee Data" in the *Proceedings of the Federal Committee on Statistical Methodology*, November 2000. [[download preprint](#)]
48. Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske "Politiques salariales et performances des entreprises : une comparaison France/Etats-Unis," *Economie et Statistique*, No. 332-333 (2000): 27-38. [Corporate Wage Policies and Performances: Comparing France with the United States] [[download preprint](#)]
49. Abowd, John M. and David Kaplan "Executive Compensation: Six Questions That Need Answering," *Journal of Economic Perspectives*, 13 (1999): 145-168. [Preprint and supplementary materials available at <http://hdl.handle.net/1813/56585>]
50. Abowd, John M., Patrick Corbel and Francis Kramarz "The Entry and Exit of Workers and the Growth of Employment: An Analysis of French Establishments" *Review of Economics and Statistics*, 81(2), (May 1999): 170-187. [[download preprint](#)]
51. Abowd, John M. and Francis Kramarz "Econometric Analysis of Linked Employer-Employee Data," *Labour Economics*, 6(March 1999): 53-74. [[download preprint](#)]
52. Abowd, John M., Hampton Finer and Francis Kramarz "Individual and Firm Heterogeneity in Compensation: An Analysis of Matched Longitudinal Employer-Employee Data for the State of Washington" in J. Haltiwanger *et al.* (eds.) *The Creation and Analysis of Employer-Employee Matched Data*, (Amsterdam: North Holland, 1999), pp. 3-24. [[download preprint](#)]
53. Abowd, John M. and Francis Kramarz "The Analysis of Labor Markets Using Matched Employer-Employee Data," in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics*, Volume 3(B), Chapter 40 (Amsterdam: North Holland, 1999), pp. 2629-2710. [[download preprint](#)]
54. Abowd, John M. Francis Kramarz and David Margolis "High Wage Workers and High Wage Firms," *Econometrica*, 67(2) (March 1999): 251-333. [[download preprint](#)]
55. Abowd, John M. Francis Kramarz, Thomas Lemieux, and David Margolis "Minimum Wages and Youth Employment in France and the United States," in D. Blanchflower and R. Freeman (eds.) *Youth Employment and Joblessness in Advanced Countries* (Chicago: University of Chicago Press, 1999), pp. 427-472. [[download](#)] [[download preprint](#)]
56. Abowd, John M. and Francis Kramarz "Internal and External Labor Markets: An Analysis of Matched Longitudinal Employer-Employee Data" in J. Haltiwanger, M. Manser, and R. Topel (eds.) *Labor Statistics and Measurement Issues* (Chicago: University of Chicago Press, 1998), pp. 357-370. [[download](#)] [[download preprint](#)]
57. Abowd, John M., Francis Kramarz, David Margolis and Kenneth Troske "The Relative Importance of Employer and Employee Effects on Compensation: A Comparison of France and the United States," in *Comparaisons internationales de salaires* (Paris: Ministère du travail et des affaires sociales and INSEE, 1996), pp. 315-327.
58. Abowd, John M. and Laurence Allain "Compensation Structure and Product Market Competition," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 207-217. [[download preprint](#)]
59. Abowd, John M., Francis Kramarz and Antoine Moreau "Product Quality and Worker Quality," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 300-322. [[download](#)]
60. Abowd, John M. and Francis Kramarz "The Microeconometrics of Human Resource Management: International Studies of Firm Practices, Introduction and Overview," *Annales d'économie et de statistique*, (January/June 1996, No. 41/42): 1-9 (French), 11-19 (English).
61. Abowd, John M. and Francis Kramarz "Les Politiques Salariales : Individus et Entreprises" (Compensation Policies: Individuals and Firms), *Revue Economique* 47 (May 1996): 611-622.

[[download preprint](#)]

62. Abowd, John M. and Francis Kramarz "The Economic Analysis of Compensation Systems: Collective and Individual" in Norman Bowes and Alex Grey, eds. *Job Creation and Loss: Analysis, Policy and Data Development* (Paris: OECD, 1996), pp. 47-54.
63. Abowd, John M. and Michael Bognanno "International Differences in Executive and Managerial Compensation" in R.B. Freeman and L. Katz, eds. *Differences and Changes in Wage Structures* (Chicago: NBER, 1995), pp. 67-103. [[download](#)]
64. Abowd, John M. and Thomas Lemieux "The Effects of Product Market Competition on Collective Bargaining Agreements: The Case of Foreign Competition in Canada," *Quarterly Journal of Economics* 108 (November 1993): 983-1014.
65. Abowd, John M. and Francis Kramarz "A Test of Negotiation and Incentive Compensation Models Using Longitudinal French Enterprise Data," in J.C. van Ours, G.A. Pfann and G. Ridder, eds. *Labour Demand and Equilibrium Wage Formation Contributions to Economic Analysis* (Amsterdam: North-Holland, 1993), pp. 111-46. [[download preprint](#)]
66. Abowd, John M. and Richard B. Freeman "Introduction and Summary" in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 1-25. [[download](#)]
67. Abowd, John M. and Thomas Lemieux "The Effects of International Competition on Collective Bargaining Outcomes: A Comparison of the United States and Canada," in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 343-67. [[download](#)]
68. Abowd, John M. "The NBER Trade and Immigration Data Files," in J.M. Abowd and R.B. Freeman, eds. *Immigration, Trade and the Labor Market* (Chicago: NBER, 1991), pp. 407-21. [[download](#)]
69. Abowd, John M. "Does Performance-based Compensation Affect Corporate Performance?" *Industrial and Labor Relations Review* 43:3 (February 1990): 525-73S. Reprinted in *Do Compensation Policies Matter?* R.G. Ehrenberg, ed. (Ithaca, NY: ILR Press, 1990), pp. 52-73.
70. Abowd, John M., George Milkovich and John Hannon "The Effects of Human Resource Management Decisions on Shareholder Value," *Industrial and Labor Relations Review* 43:3 (February 1990): 203S-236S. Reprinted in *Do Compensation Policies Matter?* R.G. Ehrenberg, ed. (Ithaca, NY: ILR Press, 1990), pp. 203-236.
71. Abowd, John M. "The Effect of Wage Bargains on the Stock Market Value of the Firm," *American Economic Review* 79:4 (September 1989): 774-800. (working paper title: "Collective Bargaining and the Division of the Value of the Enterprise.")
72. Abowd, John M. and Joseph Tracy "Market Structure, Strike Activity, and Union Wage Settlements," *Industrial Relations* 57:2 (Spring 1989): 227-50.
73. Abowd, John M. and David Card "On the Covariance Structure of Earnings and Hours Changes," *Econometrica* 57:2 (March, 1989): 411-45.
74. Vroman, Wayne and John M. Abowd "Disaggregated Wage Developments," *Brookings Papers on Economic Activity* (1:1988): 313-46.
75. Abowd, John M. and David Card "Intertemporal Labor Supply and Long Term Employment Contracts," *American Economic Review* 77:1 (March 1987): 50-68.
76. Abowd, John M. "New Development in Longitudinal Data Collection for Labor Market Analysis: Collective Bargaining Data," *American Statistical Association 1985 Proceedings of the Business and Economic Statistics Section* (Washington, DC: ASA, 1985). (invited paper)
77. Abowd, John M. and Arnold Zellner "Estimating Gross Labor Force Flows," *Journal of Business and Economic Statistics* 3 (July 1985): 254-283.
78. Abowd, John M. and Arnold Zellner "Application of Adjustment Techniques to U.S. Gross Flow Data," *Gross Flows in Labor Force Statistics*, edited by Paul Flaim and Carma Hogue, Bureau of the Census/Bureau of Labor Statistics Conference Volume (Washington, DC: GPO, 1985).
79. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the Federal and Nonfederal Sectors: Methodological Issues and Their Empirical Consequences," in *Hispanics in the U.S. Economy*, edited by G. Borjas and M. Tienda (New York: Academic Press, 1985), pp. 77-125.
80. Abowd, John M. "Economic and Statistical Analysis of Discrimination in Job Assignment," *Industrial Relations Research Association Proceedings of the Thirty-Sixth Annual Meetings* (Madison, WI: IRRRA, 1984), pp. 34-47. (invited paper)
81. Abowd, John M. and Mark Killingsworth "Do Minority/White Unemployment Differences Really Exist," *Journal of Business and Economic Statistics* 2 (January 1984): 64-72.

82. Abowd, John M. and Arnold Zellner "Estimating Gross Labor Force Flows," *American Statistical Association 1983 Proceedings of the Business and Economic Statistics Section* (Washington, DC: ASA, 1983), pp. 162-67.
83. Abowd, John M. and Mark Killingsworth "Sex Discrimination, Atrophy and the Male-Female Wage Differential," *Industrial Relations* 22 (Fall 1983): 387-402.
84. Abowd, John M. and Henry S. Farber "Job Queues and the Union Status of Workers," *Industrial and Labor Relations Review* 35 (April 1982): 354-67. [[download](#)]
85. Abowd, John M. and Orley Ashenfelter "Anticipated Unemployment, Temporary Layoffs and Compensating Wage Differentials," in *Studies in Labor Markets*, edited by S. Rosen (Chicago: University of Chicago Press for the NBER, 1981), pp. 141-170. [[download](#)]
86. Abowd, John M. "An Econometric Model of Higher Education," in *Managing Higher Education: Economic Perspectives*, A Monograph of the Center for the Management of Public and Nonprofit Enterprises (Chicago: University of Chicago Press, 1981), pp. 1-56.
87. Mulvey, Charles and John M. Abowd "Estimating the Union/Nonunion Wage Differential: A Statistical Issue," *Economica*, 47 (February 1980): 73-79.
88. Abowd, John M. and T. James Trussell "Teenage Mothers, Labor Force Participation, and Wage Rates," *Canadian Studies in Population* (1980): 33-48.

Monographs

1. Abowd, John M., Martha H. Stinson and Gary Benedetto *Final Report to the Social Security Administration on the SIPP/SSA/IRS Public Use File Project*, November 2006. [[download archival copy and Excel tables at http://hdl.handle.net/1813/43929](#)]
2. Abowd, John M. and Michael Bognanno "The Center for Advanced Human Resource Studies Managerial Compensation Database: User's Guide," March 1991.
3. Abowd, John M. and Michael Bognanno "The Center for Advanced Human Resource Studies Managerial Compensation Database: Technical Guide," March 1991.
4. Abowd, John M. *An Econometric Model of the U.S. Market for Higher Education* (New York: Garland Press, 1984).
5. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the federal and Nonfederal Sectors," in *Hispanics in the Labor Force: A Conference Report*, edited by G. Borjas and M. Tienda. Final Report to the National Employment Policy Commission (Washington, DC: GPO, 1982).
6. Abowd, John M. "Program Evaluation: New Panel Data Methods for Evaluating Training Effects," in *Program Evaluation Final Report to the U.S. Department of Labor* (Contract No. 23-17-80-01) (Washington, DC: NTIS, 1983).
7. Abowd, John M. and Mark Killingsworth "Employment, Wages, and Earnings of Hispanics in the federal and Nonfederal Sectors," in *Hispanics in the Labor Force: A Conference Report*, edited by G. Borjas and M. Tienda. Final Report to the National Employment Policy Commission (Washington, DC: GPO, 1982).
8. Abowd, John M. "Minority Unemployment, Compensating Differentials and the Effectiveness of the EEOC," in *Issues in Minority and Youth Unemployment final Report to the U.S. Department of Labor* (Contract No. 20-17-80-44) (Washington, DC: NTIS, 1982)
9. Abowd, John M. and Mark Killingsworth "Structural Models of the Effects of Minimum Wages on Employment by Age Groups," *Final Report of the Minimum Wage Study Commission*, Volume 5 (Washington, DC: GPO, 1981).
10. Abowd, John M. and Mark Killingsworth "An Analysis of Hispanic Employment, Earnings and Wages with Special Reference to Puerto Ricans," *Final Report to the U.S. Department of Labor* (Grant 21-36-78-61) (Washington, DC: NTIS, 1981).

Miscellany

1. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber, Introductory Readings in Formal Privacy for Economists (May 8, 2019, updated regularly). [[read](#), [download](#)]
2. Abowd, John M., "The Census Bureau Tries to Be a Good Data Steward in the 21st Century" International Conference on Machine Learning (ICML) 2019 [keynote address](#). [[video](#), start at minute 18:00] [[slides](#)]
3. Garfinkel, Simson L., John M. Abowd, and Christian Martindale, "Understanding Database Reconstruction Attacks on Public Data," *ACMQueue*, Vol. 16, No. 5 (September/October 2018): 28-53. [[download](#), not copyrighted]

4. Garfinkel, Simson L., John M. Abowd and Sarah Powazek "Issues Encountered Deploying Differential Privacy," *WPES'18 Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, Ontario, CA (October 2018): 133-137, DOI:10.1145/3267323.3268949. [[ArXiv preprint](#)]
5. Abowd, John M. "The U.S. Census Bureau Adopts Differential Privacy," KDD '18 Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK (August 2018): 2867, DOI:10.1145/3219819.3226070. [[download](#), subscription required], [[archival copy](#)] [[video](#)]
6. Abowd, John M., Lorenzo Alvisi, Cynthia Dwork, Sampath Kannan, Ashwin Machanavajjhala, and Jerome Reiter "Privacy-Preserving Data Analysis for Federal Statistical Agencies," Computing Community Consortium White Papers (January 2017). [[CCC white paper archive](#); [ArXiv preprint](#)]
7. Abowd, John M. "Why Statistical Agencies Need to Take Privacy-loss Budgets Seriously, and What It Means When They Do," presented to the Federal Committee on Statistical Methodology, Policy Conference, December 7-8, 2016. [[download](#)]
8. Vilhuber, Lars, John M. Abowd and Jerome P. Reiter "Synthetic Establishment Microdata around the World," *Statistical Journal of the International Association for Official Statistics*, Vol. 32 (2016): 65-68. [[download](#), open access] [[download preprint](#)]
9. Abowd, John M. "Synthetic Establishment Data: Origins and Introduction to Current Research," *Statistical Journal of the International Association for Official Statistics*, Vol. 30, No. 2 (Summer 2014): 113-115. [[download](#), subscription required] [[download preprint](#)]
10. Benedetto, Gary, Martha H. Stinson and John M. Abowd "The Creation and Use of the SIPP Synthetic Beta," U.S. Census Bureau Technical Paper (April 2013). [[download](#)]
11. Abowd, John M. and Lars Vilhuber "Science, Confidentiality, and the Public Interest," *Chance*, Vol. 24, No. 3 (Fall 2011): 58-62. [[download](#)]
12. Abowd, John M. "OnTheMap: Block-level Job Estimates Based on Longitudinally Integrated Employer-Employee Micro-data," *Association of Public Data Users Newsletter* Vol. 33, No. 2 (March/April 2010): 10-19. [[download](#)]
13. Abowd, John M. Kobbi Nissim and Chris Skinner "First Issue Editorial" *Journal of Privacy and Confidentiality*, Vol. 1, No. 1 (2009): 1-6. [[download](#)]
14. Abowd, John M. "Comments on 'Regional difference-in-differences in France using the German annexation of Alsace-Moselle in 1870-1918' by Matthieu Chemin and Etienne Wasmer" *NBER International Seminar on Macroeconomics* (2008): 306-309. [[download](#)]
15. Abowd, John M. and Julia Lane "The Economics of Data Confidentiality," *ICP Bulletin*, Volume 4, No. 2 (August 2007): 18-21. [[download preprint](#)]
16. Abowd, John M. "Rapporteur comments: International Symposium on Linked Employer-Employee Data, Econometric Issues" *Monthly Labor Review* 121:7 (July, 1998): 52-53.
17. Abowd, John M. "Discussion of 'How much do immigration and trade affect labor market outcomes' by George J. Borjas, Richard B. Freeman and Lawrence F. Katz," *Brookings Papers in Economic Activity* (1997:I): 76-82.
18. Abowd, John M. "Discussion of Gross Worker and Job Flows in Europe by M. Burda and C. Wyplosz," *European Economic Review* (1994): 1316-1320.
19. Abowd, John M. "Discussion of 'The Quality Dimension in Army Retention' by Charles Brown." in A. Meltzer (ed.) *The Carnegie-Rochester Conference on Public Policy* 33 (1990).
20. Abowd, John M. "Immigration, Trade, and Labor Markets in Australia and Canada," in *Immigration, Trade, and the Labor Market*, edited by R.B. Freeman (Cambridge, Mass: NBER, 1988), pp. 29-34.
21. Abowd, John M. "Discussion of 'Public Sector Union Growth and Bargaining Laws: A Proportional Hazards Approach with Time-Varying Treatments' by c. Ichniowski." in *Public Sector Unionism*, edited by R. Freeman (Chicago: University of Chicago Press for the NBER, 1988).
22. Abowd, John M., Ross Stolzenberg and Roseann Giarusso "Abandoning the Myth of the Modern MBA Student," *Selections The Magazine of the Graduate Management Admission Council* (Autumn 1986): 9-21.
23. Abowd, John M., Brent Moulton and Arnold Zellner "The Bayesian Regression Analysis Package: BRAP User's Manual Version 2.0," H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago, 1985.
24. Abowd, John M. and Mark R. Killingsworth "The Minimum Wage Law Winners and Losers," *The Wall Street Journal* (August 1981).

Working and Unpublished Papers

1. McKinney, Kevin L. and John M. Abowd, "Male Earnings Volatility in LEHD before, during, and after the Great Recession," (August 2020). [[download preprint](#)]
2. Abowd, John M., Gary L. Benedetto, Simson L. Garfinkel et al. "The Modernization of Statistical Disclosure Limitation at the U.S. Census Bureau," (August 2020). [[download preprint](#)]
3. Abowd, John M., Ian M. Schmutte, William Sexton, and Lars Vilhuber "Suboptimal Provision of Privacy and Statistical Accuracy When They are Public Goods," (June 2019). [[download preprint](#)]
4. Abowd, John M., Joelle Abramowitz, Margaret C. Levenstein, Kristin McCue, Dhiren Patki, Trivellore Raghunathan, Ann M. Rodgers, Matthew D. Shapiro, Nada Wasi, 2019. "Optimal Probabilistic Record Linkage: Best Practice for Linking Employers in Survey and Administrative Data," Working Papers 19-08, Center for Economic Studies, U.S. Census Bureau, handle: RePEc:cen:wpaper:19-08. [[download preprint](#)]
5. McKinney, Kevin L. Andrew Green, Lars Vilhuber, and John M. Abowd "Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On The Map" (December 2017). [[download preprint](#)]
6. Abowd, John M. and Ian Schmutte "Revisiting the Economics of Privacy: Population Statistics and Confidentiality Protection as Public Goods" (April 2017), [[download preprint](#)], published as Abowd, John M. and Ian M. Schmutte "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices," *American Economic Review*, Vol. 109, No. 1 (January 2019):171-202, DOI:10.1257/aer.20170627. [[AER](#), [ArXiv preprint](#), [Replication information](#)]
7. Abowd, John M. "Where Have All the (Good) Jobs Gone? (May 2014) Society of Labor Economists Presidential Address. [[download preprint](#)] [[accompanying audio](#)]
8. Abowd, John M., John Haltiwanger, Julia Lane, Kevin McKinney and Kristin Sandusky "Technology and Skill: An Analysis of Within and Between Firm Differences" (March 2007) NBER WP-13043. [[download preprint](#)]
9. Abowd, John M., Francis Kramarz, David N. Margolis, and Thomas Philippon "Minimum Wages and Employment in France and the United States" (February 2006). [[archival download](#)]
10. Abowd, John M., Paul Lengerhmann and Kevin L. McKinney "The Measurement of Human Capital in the U.S. Economy," (March 2003) [[download Census](#), cited on September 1, 2015] [[archival download](#)]
11. Abowd, John M., Robert Creedy and Francis Kramarz "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data," (March 2002). [[download Census](#), cited on September 1, 2015] [[archival download](#)] [[Fortran source](#)] [[Support files](#)] [[VirtualRDC archive](#)]

MAJOR GRANTS AND RESEARCH CONTRACTS

1. Associate Director for Research and Methodology and Chief Scientist U.S. Census Bureau, Intergovernmental Personnel Act (IPA) with Cornell University, June 1, 2016—March 27, 2020.
2. Research and Methodology Support Services, U.S. Census Bureau contract with Cornell University, June 1, 2015—May 31, 2016, \$268,897.
3. The Economics of Socially Efficient Privacy and Confidentiality Management for Statistical Agencies, Alfred P. Sloan Foundation awarded to Cornell University, April 1, 2015—March 31, 2019, \$535,970. (co-PIs Lars Vilhuber and Ian Schmutte)
4. RCN: Coordination of the NSF-Census Research Network, National Science Foundation SES [1237602](#) awarded to the National Institute of Statistical Sciences, July 15, 2012—June 30, 2017, transferred to Cornell University, September 2014, \$748,577. (PI Lars Vilhuber, other co-PIs Alan Karr, Jerome Reiter)
5. NCRN-MN: Cornell Census-NSF Research Node: Integrated Research Support, Training and Data Documentation, National Science Foundation Grant SES [1131848](#) awarded to Cornell University, October 1, 2011—September 30, 2016, \$2,999,614. (with William Block, Ping Li, and Lars Vilhuber)
6. A Census-Enhanced Health and Retirement Study: A Proposal to Create and Analyze an HRS Dataset Enhanced with Characteristics of Employers, Alfred P. Sloan Foundation grant awarded to the Institute for Social Research, University of Michigan with a subcontract to

- Cornell University, September 1, 2011–August 31, 2016, Cornell component \$349,608. (PI: Margaret Levenstein; other co-PIs: Matthew Shapiro, Kristin McCue and David Weir)
7. **Synthetic Data User Testing and Dissemination**, National Science Foundation Grant SES 1042181 awarded to Cornell University, September 15, 2010 to September 14, 2013, \$197,170. (Co-PI Lars Vilhuber)
 8. **CDI-Type II: Collaborative Research: Integrating Statistical and Computational Approaches to Privacy**, National Science Foundation Grant BCS 0941226 awarded to Cornell University, September 1, 2010–August 31, 2014, \$409,296. (Other PIs: Aleksandra B Slavkovic, Stephen E. Fienberg, Sofya Raskhodnikova, and Adam Smith)
 9. **TC:Large: Collaborative Research: Practical Privacy: Metrics and Methods for Protecting Record-level and Relational Data**, National Science Foundation Grant TC 1012593 awarded to Cornell University, July 15, 2010 to July 14, 2015, \$1,326,660. (Other PIs: Johannes Gehrke, Gerome Miklau, and Jerome Reiter)
 10. **The Longitudinal Employer-Household Dynamics Program**, U.S. Bureau of the Census, Interagency Personnel Act (IPA) with Cornell University, September 18, 1998 – September 17, 2000, \$260,000; renewed September 14, 2000–September 13, 2002, \$320,000; contract renewed as consultant September 14, 2002–September 13, 2003 (\$120,000); renewed as IPA September 15, 2003 – September 14, 2005 (\$384,590); renewed as IPA September 15, 2005–September 14, 2007 (\$425,215); new September 15, 2008–September 14, 2010 (497,897); renewed September 15, 2010–September 14, 2012 (532,893); continued as a contract with ACES-Research, LLC (September 17, 2012–September 16, 2013); re-established as IPA October 1, 2013–September 30, 2014 (\$231,757); re-established as IPA November 14, 2014 –May 31, 2015 (\$229,095).
 11. **Social Science Gateway to TeraGrid**, National Science Foundation Grant SES 0922005 awarded to Cornell University, July 1, 2009 to June 30, 2012, \$393,523. (Co-PI Lars Vilhuber) [Cornell Chronicle Article] [ILR News Release]
 12. **Joint NSF-Census-IRS Workshop on Synthetic Data and Confidentiality Protection**, July 2009 Washington, DC, National Science Foundation Grant SES 0922494 awarded to Cornell University, July 1, 2009 to June 30, 2010, \$18,480. (Co-PIs Lars Vilhuber, Jerome Reiter, and Ron Jarmin)
 13. **The Economics of Mass Layoffs: Displaced Workers, Displacing Firms, Causes and Consequences**, National Science Foundation Grant SES-0820349 awarded to Cornell University, October 1, 2008 to September 30, 2010, \$245,950. (Co-PI Lars Vilhuber)
 14. **LEHD Developmental and Confidentiality Research**, Census Bureau Contract to Abt Associates with subcontract awarded to Cornell University, August 1, 2007 to September 30, 2008, \$358,270.
 15. **CT-T: Collaborative Research: Preserving Utility While Ensuring Privacy for Linked Data**, National Science Foundation Grant CNS-0627680 awarded to Cornell University, September 5, 2006 to August 31, 2009, \$488,950. (PI Johannes Gehrke)
 16. **LEHD Confidentiality Research**, Census Bureau Contract to Abt Associates with subcontract awarded to Cornell University, October 1, 2004 to September 30, 2005, \$230,155.
 17. **ITR-(ECS+ASE)-(dmc+int): Info Tech Challenges for Secure Access to Confidential Social Science Data**, National Science Foundation Grant SES-0427889 awarded to Cornell University, October 1, 2004 to September 30, 2007, \$2,938,000. (Co-PIs Matthew D. Shapiro, Ronald Jarmin, Stephen F. Roehrig, and Trivellore Raghunathan) [Cornell Chronicle article]
 18. **EITM: Developing the Tools to Understand Human Performance: An Empirical Infrastructure to Foster Research Collaboration**, National Science Foundation Grant SES-0339191 awarded to Cornell University, October 1, 2004 to September 30, 2007, \$337,455 (Co-PIs John Haltiwanger and Ron Jarmin)
 19. **The New York Research Data Center**, National Science Foundation Grant SES-0322902 awarded to the NBER, August 1, 2003 to July 31, 2004, \$300,000. (PI Neil G. Bennett, Other co-PIs Bart Hobijn, Erica L. Groshen, Robert E. Lipsey)
 20. **Workshop on Confidentiality Research**, National Science Foundation Grant SES-0328395 awarded to the Urban Institute, June 1, 2003 – May 31, 2004, \$43,602. (Co-PI Julia Lane)
 21. **Firms, Workers and Workforce Quality: Implications for Earnings Inequality and Economic Growth**, Alfred P. Sloan Foundation Grant 22319-000-00 awarded to the Urban Institute, January 2003–January 2006, \$1,400,000. (Co-PIs John Haltiwanger, Julia Lane, J. Bradford Jensen, Fredrick Knickerbocker, and Ronald Prevost)
 22. **The Demand for Older Workers: Using Linked Employer-Employee Data for Aging Research**, National Institute on Aging, R01-AG18854-01 to Cornell University, July 1, 2002 – April 30,

- 2007, \$1,753,637. (Co-PIs John Haltiwanger, Andrew Hildreth, and Julia Lane)
23. Workers and Firms in the Low-wage Labor Market: Interactions and Long Run Dynamics, Russell Sage Foundation, Rockefeller Foundation, and Department of Health and Human Services (ASPE) to the Urban Institute \$700,000, September 1, 2001 August 31, 2003. (Co-PIs John Haltiwanger, Harry Holzer, and Julia Lane)
24. From Workshop Floor to Workforce Clusters: A New View of the Firm, Alfred P. Sloan Foundation, 99-12-12 to the Urban Institute, March 1, 2000 – March 31, 2002, \$314,604. (Co-PIs John Haltiwanger and Julia Lane)
25. [Dynamic Employer-Household Data and the Social Data Infrastructure](#), National Science Foundation, SES-9978093 to Cornell University, September 28, 1999 – September 27, 2005, \$4,084,634. (Co-PIs John Haltiwanger and Julia Lane)
26. The Longitudinal Employer-Household Dynamics Program, National Institute on Aging, interagency funding to the United States Census Bureau, September, 1999 – August, 2001, \$490,000. Renewed September 2001– August 2004, \$750,000 (Co-PIs John Haltiwanger and Julia Lane) [[Cornell Chronicle article](#)]
27. [Individual and Firm Heterogeneity in Labor Markets: Studies of Matched Employee-Employer Data](#), National Science Foundation SBR 9618111 to the NBER, March 15, 1997 – February 28, 2002, \$243,361.
28. Creation of an Employer Identification Link File and Addition of Employer Information to the National Longitudinal Survey of Youth 1979 Cohort, Bureau of Labor Statistics (subcontracted by NORC, University of Chicago, Chicago, IL 60637), July 1, 1995 – December 31, 1997, \$82,946.
29. [Employment and Compensation Policies: Studies of American and French Labor Markets Using Matched Employer-Employee Data](#), National Science Foundation SBR 9321053 to the NBER, July 1, 1994 – June 31, 1997, \$ 185,257. (Co-PIs David Margolis and Kenneth Troske)
30. [Compensation System Design, Employment and Firm Performance: An Analysis of French Microdata and a Comparison to the United States](#), National Science Foundation, SBR 9111186 to Cornell University, July 1, 1991 – December 30, 1994, \$174,565.
31. [The Effects of Collective Bargaining and Threats of Unionization on Firm Investment Policy, Return on Investment, and Stock Valuation](#), National Science Foundation, SES 8813847 to the NBER, July 1, 1988 – June 30, 1990, \$81,107.
32. [Improving the Scientific Research Utility of Labor Force Gross Flow Data](#), National Science Foundation, SES 85-13700 to the NBER, April 15, 1986 – March 31, 1988, \$69,993.
33. Program Evaluation: New Panel Data Methods for Evaluating Training Effects, U.S. Department of Labor Contract 23-17-80-01 to NORC at the University of Chicago, 1983.
34. Minority Unemployment, Compensating Differentials and the Effectiveness of the EEOC, U.S. Department of Labor Contract 20-17-80-44 to NORC at the University of Chicago, 1982.
35. An Analysis of Hispanic Employment, Earnings and Wages with Special Reference to Puerto Ricans, U.S. Department of Labor Grant 21-36-78-61, 1981.

PROFESSIONAL SERVICE, SURVEYS, AND DATA COLLECTION

1. [Canadian Research Data Centre Network Inaugural Board](#) 2017-2019.
2. American Economic Association, Committee on Economic Statistics (AEAWeb) 2013-2018.
3. National Academy of Sciences, Committee on National Statistics (CNSTAT) 2010-2013; reappointed 2013-2016.
4. National Academy of Sciences, CNSTAT, Panel on Measuring and Collecting Pay Information from U.S. Employers by Gender, Race, and National Origin, (Chair) 2011-2012.
5. National Academy of Sciences, CNSTAT, Panel on Measuring Business Formation, Dynamics and Performance, 2004-2007.
6. National Academy of Sciences, CNSTAT, Panel on Data Access for Research Purposes, 2002-2005.
7. Executive Committee, Conference on Research in Income and Wealth 2002-.
8. Distinguished Senior Research Fellow, LEHD Program, U.S. Census Bureau 1998-2016.
9. Social Science and Humanities Research Council (Canada), Major Collaborative Research Initiatives review panel, 1997, 1998.
10. Technical Advisory Board for the National Longitudinal Surveys of the Bureau of Labor Statistics, 1988-1990, 1992-2001, Chair 1999-2001.
11. National Science Foundation, Economics Panel, 1990-91, 1992-93; KDI Panel 1999; Infrastructure Panel 2000; CDI Panel 2008; CDI Panel 2009.

12. Principal Investigator for The Center for Advanced Human Resource Studies Managerial Compensation Data Base. sponsored by the Cornell University Center for Advanced Human Resource Studies, 1989-1994.
13. Principal Investigator for A Longitudinal Data Base of Collective Bargaining Agreements. Sponsored by the Bureau of National Affairs and the University of Chicago Graduate School of Business, 1985.

PROFESSIONAL ORGANIZATIONS

1. American Economic Association
2. American Statistical Association
3. Econometric Society
4. Society of Labor Economists
5. International Statistical Institute
6. International Association for Official Statistics
7. National Association for Business Economics
8. American Association of Wine Economists
9. American Association for Public Opinion Research
10. Association for Computing Machinery
11. American Association for the Advancement of Science

PERSONAL INFORMATION

United States citizen

Personal email: john.abowd@gmail.com

APPENDIX B — 2010 RECONSTRUCTION-ABETTED RE-IDENTIFICATION SIMULATED ATTACK

1. This appendix provides a high-level summary of the reconstruction-abetted re-identification attack simulation that the Census Bureau conducted on the released 2010 Census data. To assess the risk of a reconstruction-abetted re-identification attack, the Census Bureau conducted a series of statistical exercises to quantify the contemporaneous and future risk that individual responses could be disclosed. The Census Bureau has completed two simulated attacks that address the re-identification risk of a 100% microdata file (a file with detailed, individual-level records for every person enumerated in the census) reconstructed from the published Summary File 1 data. The 2010 Summary File 1, usually called SF1, includes the 2010 P.L. 94-171 Redistricting Data Summary File, the 2010 Advanced Group Quarters Data Summary File, and the bulk of the demographic and housing characteristics released from the 2010 Census in tabular format.¹ The fundamental structure of these simulations is as follows.

SIMULATED RECONSTRUCTION ATTACK

2. Database reconstruction is the process of statistically re-creating the individual-level records from which a set of published tabulations was originally calculated. That is, database reconstruction attempts to “reverse engineer” the confidential input data used in a statistical tabulation system.
3. The Census Bureau released over 150 billion statistics as part of the 2010 Census. The simulated reconstruction attack used as its input a small fraction of those statistics—approximately 6.2 billion statistics contained in the following published SF1 tables from the 2010 Census:

P001 (Total Population by Block)
 P006 (Total Races Tallied by Block)
 P007 (Hispanic or Latino Origin by Race by Block)

¹ See the technical documents in [Summary File 1 Dataset \(census.gov\)](https://www.census.gov/data/tables/2010/sf1.html).

P009 (Hispanic or Latino, and Not Hispanic or Latino by Race by Block)
P011 (Hispanic or Latino, and Not Hispanic or Latino by Race for the Population 18 Years and Over by Block)
P012 (Sex by Age by Block)
P012A-I (Sex by Age by Block, iterated by Race)
P014 (Sex by Single-year-of-age for the Population under 20 Years by Block)
PCT012A-N (Sex by Single-year-of-age by Tract, iterated by Race)

4. The reconstruction of the 2010 Census microdata for the sex, age, race, Hispanic/Latino ethnicity, and census block variables was carried out by constructing a system of equations consistent with the published tables listed above that, once solved, could then be converted into microdata. This system of equations was solved using commercial mixed-integer linear programming software (Gurobi).
5. Because the parameters of the 2010 Census swapping methodology included invariants on total population and voting age population at the block level, the reconstruction was able to exactly reconstruct all 308,745,538 million records with correct block location and voting age (18+). Then, leveraging the race (63 categories), Hispanic/Latino origin, sex, and age (in years) data from the specified tables, the simulated attack was able to further reconstruct those variables on the individual-level records.
6. To assess the accuracy of these reconstructed individual-level records, the team performed exact record linkage of the five variables in the reconstructed microdata to the same five variables in the Census Edited File (CEF, the confidential data) and Hundred-percent Detail File (HDF, the confidential swapped individual-level data before tabulation). The results are summarized in Table 1. The “left” file of the record linkage is in the first column. The “right” file is the reconstructed microdata from SF1.

Table 1 Agreement Rates between the Reconstructed Microdata and the 2010 Census Edited File and Hundred-percent Detail File					
	Record Counts		Agreement Rates		
Left file	In Left	In Reconstructed	Exact	Fuzzy Age	One error
CEF	308,745,538	308,745,538	46.48%	70.98%	78.31%
HDF	308,745,538	308,745,538	48.34%	73.33%	80.39%

DRB clearance number CBDRB-FY21-DSEP-003

7. The agreement rates shown in Table 1 include block (which was never wrong), sex, age (in years), race (63 OMB categories), and Hispanic ethnicity and are computed as a percentage of the total population. Exact agreement means all five variables agreed precisely bit for bit. Fuzzy-age agreement means that block, sex, race, and Hispanic ethnicity agreed exactly, but age agreed only ± 1 year (e.g., age 25 on the CEF is in fuzzy-age agreement with ages 24, 25, and 26 on the reconstructed data). The one-error agreement rate allows one variable — sex, age (outside ± 1 year), race or ethnicity to be wrong.
8. Most errors in the reconstructed file are that the age variable is off by ± 2 years rather than ± 1 year. This error is the balance of the width of the 5-year categories used in the block-level summaries. Hence, even though the disclosure avoidance requirement for the 2010 Census SF1 tabular summaries specified block-level aggregation to 5-year bins for those age 20 and over, the effective aggregation was far less.
9. Figure 1 shows the distribution of agreement rates by block size. Agreement rates are only substantially lower than the population averages shown in Table 1 for blocks with populations between 0 and 9 people, which is where the Census Bureau has said it concentrated the swaps.² However, uniqueness on sex, age, race, and ethnicity is not limited to small population blocks. *This is one of the principal failures of the 2010 tabular disclosure avoidance methodology — swapping provided protection for households deemed “at risk,” primarily those in blocks with small populations, whereas for the entire 2010 Census a full 57% of the persons are population uniques on the basis of block, sex,*

² McKenna, L. (2018), “Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing,” <https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf>, p. 8.

age (in years), race (OMB 63 categories), and ethnicity. Furthermore, 44% are population uniques on block, age and sex.³

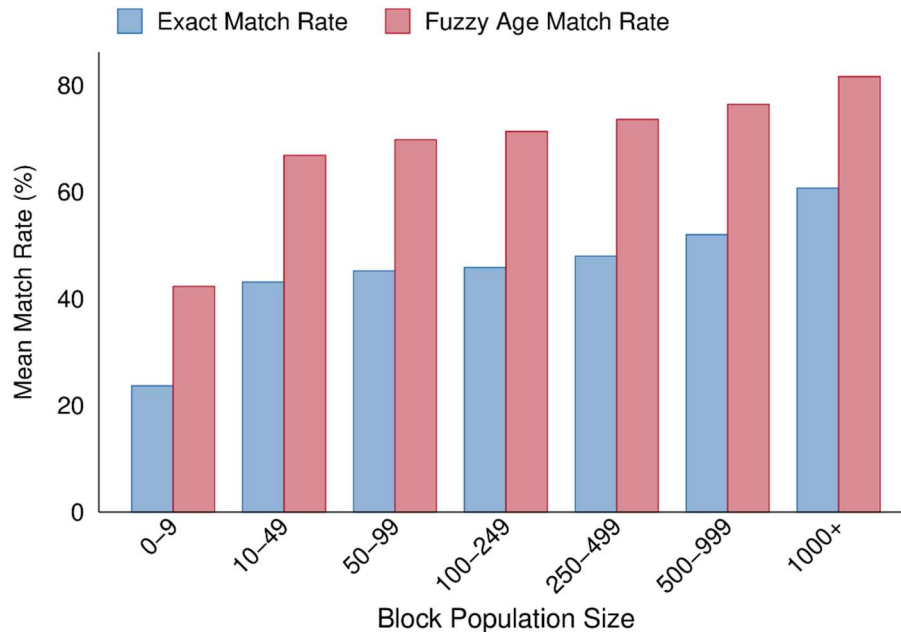


Figure 1 Block-level agreement rates between the reconstructed 2010 Census micro-data and the 2010 Census Edited File by population in the block
DRB clearance number CBDRB-FY21-DSEP-003.

10. Although there are no recent re-identification studies for decennial Public Use Micro-data Samples (PUMS) with geography coded to the Public Use Microdata Area (PUMA), the Census Bureau continues to use 100,000 persons as the minimum population threshold for such areas and has coded geography on the 2010 PUMS and all American Community Survey (ACS) PUMS using these PUMAs. Since sex and age (single years) are population uniques at the tract level for only 0.18% of persons, this may still be justifiable for a 10% sample of 2010 Census records, but the potential re-

³ The statistics in this paragraph are cleared for public release by the Census Bureau Disclosure Review Board (CBDRB-FY21-DSEP-003).

identification rate for a 100% public-use microdata file geocoded to the block level is certainly quite large.

11. The reconstruction experiment demonstrated that existing technology can convert the Census Bureau's traditional tabular summaries of Census data which was released in 2010 into a 100% coverage microdata file geocoded to the block level with very limited noise which was not released in 2010. This microdata file contains so much detail that it would have been deemed "unreleasable" if it had been proposed in conjunction with the original 2010 Census data products.
12. The ability to reconstruct the microdata means that there is now a significant disclosure risk for the 2010 Census Summary Files 1 and 2 (SF1, SF2) and the American Indian Alaska Native Summary File (AIANSF) data. There are approximately 150 billion statistics in the SF1, SF2, and AIANSF summaries (recall that the 2010 P.L. 94-171 Redistricting Data Summary File and the 2010 Advanced Groups Quarters Summary File are part of SF1). Because of the features noted above, releasing this many very accurate statistics made the ensemble of those publications equivalent to releasing the 2010 Hundred-percent Detail File (HDF), the swapped version of and the 2010 Census Edited File (CEF). There can be no uncertainty about this: *the 2010 Census tabular publications were equivalent to releasing every tabulation variable in the 2010 HDF in universe public-use microdata files without the hierarchical structure--person and household records can be fully reconstructed, but not directly linked to each other.* The team that demonstrated this vulnerability stopped after reconstructing person-level records for block, sex, age (in years), race (63 OMB categories), and Hispanic ethnicity because the vulnerability had been fully exposed mathematically and demonstrated empirically.
13. There are 308,745,538 (U.S. only) person records and 131,704,730 housing unit records in both the 2010 HDF and CEF, linked in their correct hierarchy. For the unswapped records in HDF, the images are identical to their CEF counterparts. For the swapped household records, the block identifier, household size, adult (age 18+) household

size, occupancy, and tenure variables are identical to their unswapped counterparts and on the person record the voting-age variable is identical to the unswapped counterpart.

14. As the documentation in McKenna (2018, 2019a) makes clear, a public-use microdata file containing the 308,745,538 person records in the HDF including only the five tabulation variables block, sex, age (in years), race (63 OMB categories), and Hispanic ethnicity is so disclosive that it would not have passed the disclosure avoidance criteria used for the 2010 Census Public-Use Microdata Sample.⁴ Furthermore, the same file would not have passed the disclosure avoidance criteria applied to SF1 itself.⁵ The official 2010 PUMS had a geographic population threshold of 100,000, collapsed categories to national population thresholds of 10,000, used partially synthetic data for the group quarters population, and “topcoding, bottom-coding, and noise infusion for large households.” The PUMS was sampled from the swapped version of the 2010 HDF, not the Census Edited File.

15. The additional disclosure avoidance methods used for the 2010 PUMS are explicitly noted on pages 2-1 and 2-2 of its technical documentation. The definition of a Public Use Microdata Area also explicitly references its confidentiality protection purpose:

“The Public Use Microdata Sample (PUMS) files contain geographic units known as Public Use Microdata Areas (PUMAs). To maintain the confidentiality of the PUMS data, a minimum population threshold of 100,000 is set for PUMAs. Each state is separately identified and may be comprised of one or more PUMAs. PUMAs do not cross state lines. (page 1-2, emphasis added)”

⁴ McKenna, L. (2019a) “Disclosure Avoidance Techniques Used for the 1960 Through 2010 Decennial Censuses of Population and Housing Public Use Microdata Samples,” Research and Methodology Technical Report available at [Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census](#).

⁵ McKenna, L. (2018)

16. This failure to apply microdata disclosure avoidance matters because the reconstructed 2010 microdata for block, sex, age (in years), race (63 OMB categories), and Hispanic ethnicity are a very accurate image of the HDF, and the HDF is a very accurate image of the CEF, which is the reason that it is also confidential. Consequently, the new technology-enabled possibility of accurately re-constructing HDF microdata from the published tabular summaries and the fact that those reconstructed data do not meet the disclosure avoidance standards established at the time for microdata products derived from the HDF demonstrate that the swapping methodology as implemented for the 2010 Census no longer meets the acceptable disclosure risk standards established when that swapping mechanism was selected for the 2010 Census.
17. Having demonstrated that a 100% microdata file can be successfully reconstructed from the published 2010 Census tabulations, the Census Bureau proceeded to use these reconstructed microdata to simulate a re-identification attack on those data.

DE-IDENTIFICATION ATTACK SIMULATION

18. The simulated re-identification attack proceeds as follows. Identify a person-level data source file that contains name, address, sex, and birthdate (e.g., commercially available data). Convert the names and addresses to their corresponding Census Bureau Protected Identification Key (PIK). Identify the corresponding census block for every address in the source file. Then, looping through all the records in the reconstructed microdata file produced from the reconstruction, find the first record in the source file that matches exactly on block, sex, and age. Once this step is completed, run through the remaining unmatched records from the reconstructed microdata and find the first unmatched record from the source file that matches exactly on block and sex, and matches on age plus or minus 1 year.
19. When both steps have been completed, output the records with successful matches from these two passes. These are called *putative re-identifications* because they appear

to link the reconstructed microdata to a real name and address associated with the block, sex, age, race, and ethnicity on the reconstructed microdata. These are the records the hypothetical attacker thinks are re-identified.

20. Putative re-identifications are not necessarily correct. An external attacker would have to do extra field work to estimate the *confirmation rate* – the percentage of putative re-identifications that are correct. An external attacker might estimate the confirmation rate by contacting a sample of the putative re-identifications to confirm the name and address. An external attacker might also perform more sophisticated verification using multiple source files to select the name and address most consistent with all source files and the reconstructed microdata.
21. At the Census Bureau we usually estimate the confirmation rate as a percentage of the total population, not as a percentage of the putative re-identifications, by performing a similar record linkage exercise of the putative re-identifications against the CEF, looking for exact matches on all variables (including PIK, block, sex, age, race, and ethnicity), followed by a second pass looking for exact matches except age, which is allowed to vary by plus or minus 1 year. Once these two passes have been completed, the matched records are the confirmed re-identifications, using exact match on PIK, block, sex, race (63 OMB categories), and ethnicity and match on age +/- 1 year as the definition of correct. The remaining unmatched records from the putative re-identifications of the reconstructed data are the unconfirmed re-identifications.
22. Table 2 shows the results of two such re-identification confirmation exercises. The first of these uses the combined commercial databases from Experian Marketing Solutions Incorporated, Infogroup Incorporated, Melissa Data Corporation, Targus Information Corporation, and VSGI LLC as the source file for name, address, sex, and age. This exercise simulates data quality circa 2010 for an external attacker relying on the consumer information in these databases. These results are in the row labeled “Commercial.” This re-identification experiment was the basis for the statistics released at the

American Association for the Advancement of Science 2019 annual meeting. Putative re-identifications were 138 million (45% of the 2010 Census resident population of the U.S.). Confirmed re-identifications were 52 million (17% of the same population).

23. Using the commercial data as the source for name, address, sex, and age is, as discussed in the main declaration, a best-case assumption. We know that these data exist and were available circa 2010 because that is when the Census Bureau acquired them. An external attacker, using the versions that the Census Bureau acquired and the relatively straightforward methodology above, would succeed at least as often as we did. This means that at least 52 million persons enumerated during the 2010 Census could be correctly re-identified using the attack strategy outlined here.
24. Suppose the external attacker had name, address, sex, and age of much better quality than the five commercial sources above. How much better could that attacker do using exactly the same strategy? This question can be answered by substituting the name, address, sex, and age from the 2010 CEF as the source file in the putative re-identification simulation. This is not cheating because no extra information in the CEF such as race, ethnicity or household structure is used for the source file. Hence, it is a proper worst-case scenario, and the one historically used by the Census Bureau in assessing microdata re-identification risk (see McKenna 2019b). If the external data on name, address, sex, and age are comparable to the 2010 Census, then the attacker will putatively re-identify 238 million persons (77% of the 2010 Census resident U.S. population). Confirmed re-identifications will be 179 million (58% of the same population). This means that with the best quality external data, relative to the 2010 Census, as many as 179 million persons could be correctly re-identified using the attack strategy outlined here.

Table 2 Record Linkage Summary from Commercial and CEF Record Sources				
PIK, Block, Age, Sex Record Linkage Source	Available Records	Records with PIK, Block, Sex, and Age	Putative Re-identifications using Source	Confirmed Re-identifications
Commercial	413,137,184	286,671,152	137,709,807	52,038,366
CEF	308,745,538	279,179,329	238,175,305	178,958,726
DRB clearance number CBDRB-FY21-DSEP-003.				

25. The record linkage results reported in Table 2 can be interpreted using two additional statistical quality measures: the *recall rate* and the *precision rate*. Taken together, these measures assess how successful an attacker can be at re-identifying records and how confident the attacker would be in those re-identifications.
26. *Recall rate*. The recall rate is the percentage of available source records that are correctly re-identified. Its numerator is the same as the confirmation rate, but its denominator is the number of records in the source file with sufficient information to perform the putative re-identification record linkage. For the two source files analyzed in these experiments, Table 2 shows the denominators for the recall rate in the column “Records with PIK, Block, Sex, and Age,” which gives the count of records with sufficient information to generate a putative match. Table 3 shows the recall rates for the two experiments. Both are greater than the respective confirmation rate because both the commercial data and the CEF have fewer usable records than the U.S. resident population. A critical result is the recall rate of 64% when the CEF is used as the source file. This result means that an external attacker with high quality name, address, sex, and age information succeeds in re-identification almost two times in three.

Table 3 Confirmation and Recall Rates		
Source	Percentage of U.S. Resident Popula- tion (Confirmation Rate)	Percentage of Complete Data Population (Recall Rate)
Commercial	16.85%	18.15%
CEF	57.96%	64.10%
DRB clearance number CBDRB-FY21-DSEP-003.		

27. *Precision rate.* Precision is the ratio of confirmed to putative re-identifications. It answers the question “How often is the attacker’s claimed re-identification correct as a percentage of the names the attacker attached to reconstructed census microdata?” Table 4 summarizes the precision rates for the two experiments. The precision of the experiment reported in February 2019 was 38% (first row of Table 4). The precision of the worst-case experiment is 75% (second row of Table 4). *This result means that an attacker using high-quality name, address, sex, and age data is correct three times out four.*

Table 4 Precision Rates	
Source	Confirmed Percent- age of Putative Re- identification (Peci- sion Rate)
Commercial	37.79%
CEF	75.14%
DRB Clearance number CBDRB-FY21-DSEP-003.	

28. To be successful, an attacker does not have to be a commercial entity, nor does a successful attack need to use commercially available data. Many agencies of federal, state and local governments in the U.S. now possess high-quality data on name, address,

sex, and age. When preparing public-use microdata files that contain variables that other agencies can access exactly, it has long been the practice to coarsen such data to prevent non-statistical uses by other agencies (see McKenna 2019b). Applying such precautions to decennial census data products would imply severe limitations on the variables published at the block level, even in the presence of swapping.

29. In conclusion, the Census Bureau's simulated reconstruction-abetted re-identification attack definitively established that the tabular summaries from the 2010 Census could be used to reconstruct individual record-level data containing the tabulation variables with their most granular definitions. Such microdata violated the disclosure avoidance rules that the Data Stewardship Executive Policy Committee had established for the 2010 Census and would not have been released had they been proposed as an official product because they posed too great a disclosure risk. The disclosure risk presumed by the 2010 standards recognized the excessive risk of re-identification if block geographic identifiers were placed on a 100% enumeration microdata file along with age (in years) and sex. The Census Bureau believed in 2010 that the minimum population that the geographic identifier could represent in such microdata is 100,000 persons – the size of a Public-Use Microdata Area. That belief was strongly confirmed by the simulated re-identification attack. Somewhere between 52 and 179 million person who responded to the 2010 Census can be correctly re-identified from the re-constructed microdata.

FINAL 2/16/10**DSEP Meeting Record**

Topic: Updates

Meeting Date: 1/14/10

Attendees:

<i>Position</i>	<i>Attending for Position</i>
Deputy Director (Chair)	Tom Mesenbourg
AD, Administration	Ted Johnson
AD, Decennial	David Whitford
AD, Demographic	Cheryl Landman
AD, Economic	Harvey Monk
AD, Field	Marilia Matos
AD, IT	Brian McGrath
AD, Strategic Planning	Nancy Gordon
Rep. for Statistical Methodology	Tommy Wright
Senior Advisor for Data Management	Teresa Angueira
Chief, ITSO	
Chief, OAES	Kathleen Styles
Chief Privacy Officer	Mary Frazier
Also Attending:	Carol Comisarow, Ron Jarmin, David Raglin, Sharon Stern, Laura Zayatz, Michael Hawes

UPDATES

Background

[REDACTED]

Disclosure Review Board

- The DRB has been using enhanced disclosure avoidance procedures and methods for projects involving sensitive topics and/or sensitive populations. These procedures were implemented in response to an August 2004 memo from Director Kincannon. Though the memo was superseded by the Custom Tabulations policy, the DRB was not informed of this, and has not changed its procedures for sensitive topics and populations.
- Laura Zayatz also voiced the DRB's concern about planning for the 2020 Census and continuing to release data at the block level, as block populations continue to decrease (e.g. 40% of blocks in North Dakota have only 1 household in them)

[REDACTED]

Action Items

1. The DRB will develop recommendations for addressing the issue of disclosure review for sensitive populations. They will present their recommendations to DSEP once they have been vetted at the staff level.
2. DSEP agrees that the problem of block population size and disclosure avoidance is real, and that it deserves attention in the context of 2020 planning.

[REDACTED]

DSEP Meeting Record

Topics:

Public Use File Reidentification Threats Update

Meeting Date: February 5, 2015

Attendees:

<i>Position</i>	<i>Attending for Position</i>
Deputy Director (Chair)	<i>absent</i>
AD, Administration	<i>absent</i>
AD, Decennial	Lisa Blumerman
AD, Demographic	Enrique Lamas
AD, Economic	<i>absent</i>
AD, Field	Jay Keller
AD, IT	Avi Bender
AD, Research and Methodology	Tom Louis
AD, 2020 Census	Lisa Blumerman
AD, Communications	Kim Collier
AD, Performance Improvement	Susan Reeves
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Chief Information Security Officer	Tim Ruland
Asst. Director, Research and Methodology	Ron Jarmin
Also Attending:	Barbara Downs, Randy Becker, Byron Crenshaw, Laura McKenna, Heather Madray, Raj Dwivedy, Julie Atwell, Mike Castro

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]




[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Public Use File Reidentification Threats Update

Background and Discussion:

On December 11, 2014, DSEP discussed a reidentification issue that occurred involving a Public Use File (PUF) produced as part of the New York City Housing Vacancy Survey (NYCHVS). At that meeting, DSEP commissioned a team to pursue the recommendations presented to DSEP in July 2014 in the paper titled “PUMS File Re-identification Threats and Potential Solutions for Mitigating those Threats.”

After discussing the logistics with some key stakeholders, and the difficulties of managing so many different angles on one team, DSEP approved a two-pronged approach to pursuing the paper’s recommendations.

The Center for Disclosure Avoidance Research (CDAR) has recently received authorization to hire new staff to focus primarily on synthetic data and reidentification research. This group is preparing a research proposal that focuses on the disclosure avoidance side of the PUF reidentification issue.

In addition to these efforts, the Demographic Programs Directorate (ADDP) will charter a team that focuses on the broader future of PUFs as well as some of the non-technical means of disclosure avoidance discussed in the July 2014 paper. This team will discuss Terms of Use for PUFs, restricted access, and other methods. This team will have representation from all of the impacted directorates. DSEP also recommended the team engage with external researchers on some of these ideas, and address their concerns.

Action Items:

- CDAR will prepare a research proposal to outline future Census Bureau efforts in Synthetic Data and Reidentification Research.
- ADDP will charter a team to evaluate the future of PUFs and explore some of the non-technical solutions outlined in the July 2014 paper.

DSEP Meeting Record

Topics: Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing (John Abowd, ADRM)

Record-level Re-identification Linkages for Evaluating the 2010 and 2020 Census Disclosure Avoidance Systems (John Abowd, ADRM)



Meeting Date: May 10, 2017

<i>Position</i>	<i>Attending for Position</i>
Deputy Director (Chair)	Laura Furgione
CAO	David Ziaya
CFO	Joanne Crane
AD, Decennial	Lisa Blumerman
AD, Demographic	Karen Battle
AD, Economic	Ron Jarmin
AD, Field	Joan Hill
AD, IT	Nitin Naik
AD, Research and Methodology	John Abowd
AD, 2020 Census	Lisa Blumerman
AD, Communications	Stephen Buckner
AD, Performance Improvement	Ted Johnson
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Senior Advisor Designee from the Director's Office	<i>absent</i>
Chief Information Security Officer	<i>absent</i>

Asst. Director, Research and Methodology	John Eltinge
Also Attending:	Simson Garfinkel, Byron Crenshaw, Eloise Parker, Ashley Landreth, Mike Castro, Harold Saintelien, Janean Darden, Julie Atwell

Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing

Background:

The Census Bureau's Research and Methodology Directorate (ADRM) is researching and developing disclosure avoidance methods and systems to replace those used for Census 2000 and the 2010 that were not designed to protect against database reconstruction attacks. ADRM is establishing the 2020 Disclosure Avoidance System (DAS), a formally private system based on the theoretical model known as differential privacy. This is the available technology for controlling reconstruction attacks.

The 2020 DAS team is working to establish adjustable formal privacy parameters for the 2018 End-to-End test. They are seeking DSEP concurrence with the Disclosure Review Board's (DRB's) April 10, 2017 determination that six data elements of PL 94-171 can continue to be published as enumerated. The team will test methods and systems with these elements published as enumerated for the 2018 End-to-End with the goal of making sound recommendations to DSEP for the full 2020 DAS. These elements to be published as enumerated are:

- the number of occupied housing units per block,
- the number of vacant housing units per block,
- the number of households per block,
- the number of adults (age 18+) per block (where the definition of an adult is inferred from the structure of the PL94-171 age categories),
- the number of children (age less than 18) per block (where the definition of a child is also inferred from the structure of the PL94-171 age categories),
- and the number of persons per block.

ADRM expects to perform follow-up analyses of the test products developed for the End-to-End Test. Because there is no national sample in 2018, some aspects of the differentially private system cannot be implemented in the End-to-End Test. They will have to be simulated from the 2010 Census data. This means that the demonstration data from the test can be made as noisy as DSEP wishes. However, there is only time to implement algorithms that maintain confidentiality with the six data elements in the 2010 PL94-171 redistricting data. There will be both policy and disclosure avoidance issues surrounding how broadly those products can be disseminated. Those issues will be brought to the DRB in a timely fashion.

ADRM also notes that DSEP will be asked to assume a formal policy consultant role for setting the confidentiality protection parameters for the final 2018 End-to-End Test and the 2020 DAS. The charter for DSEP currently delegates the authority to set disclosure avoidance standards to the DRB, with review by DSEP if necessary. However, these parameters now must be public in a formal privacy system. Furthermore, they, like any other operational decision need to be

discussed and set in a manner consistent with their importance in the publication of results from the 2020 Census. The privacy-loss setting recommended by DRB and DSEP, and accepted by the Director, will be implemented in the production system.

Requests to DSEP:

Request 1: Concurrence with the DRB's decision on the PL 94-171 file items that can be published as enumerated.

In order to meet the timeline for the 2018 End-to-End Test, the version of the DAS under development for the test is limited in scope to the PL94-171 redistricting data. ADRM will not have time to experiment with a suite of potential implementations. And, in particular, ADRM will not have time to modify certain implementation decisions. They will be put back on the table for the full 2020 Disclosure Avoidance System and the decision on these six specific items may be revisited.

Request 2: Concurrence with Change to DRB Operating Principles Related to 2020 Census

The second request is for DSEP concurrence on a change in the operating principles of the DRB for issues related to disclosure avoidance in the 2020 Census of Population. Because the differentially private disclosure avoidance methods operate on the ensemble of proposed publications, DSEP is asked to concur that any disclosure avoidance request for publications from 2020 Census data be routed to the 2020 DAS team first. Those requests should not be considered by the DRB until the 2020 DAS team supplies a memo stating that the requested publication can or cannot be incorporated into the total privacy-loss accounting.

This is not a request for a moratorium on approvals for decennial data releases or design. The privacy-loss budget itself and its allocation to various components of the publication system are policy decisions that the 2020 Disclosure Avoidance System team will not make. Those decisions will ultimately be made in a manner consistent with the charters of the DRB and DSEP, and defended by the Director.

There is very little historical guidance for this process. We need to develop practical use cases that illustrate the consequences of publication decisions under alternative privacy-loss scenarios. We need to document the extent to which a best-effort reconstruction of the 2010 Hundred-percent Detail File (HDF) is correlated with the actual HDF. This is going to take some time. In the interim, ADRM is asking the DRB to take a leadership role in making these important choices by enabling the development of technologies better adapted to global risk management.

Discussion:

DSEP recognized the value in ADRM's efforts to assemble a skilled team of experts in an effort to modernize Census Bureau disclosure avoidance techniques using formal privacy methods.

This is essential in light of research that demonstrates that we must protect against database reconstructions that could lead to re-identification.

DSEP discussed the details of the six data elements from PL 94-171 and considered the necessity of including all of these in the proposed 2020 DAS research. ADRM requested that all elements remain available for the 2018 test research with a reconsideration for the full 2020 DAS, once the Census Bureau understand the outcomes. Conversations with the Department of Justice for Voting Rights Acts requirements with PL 94-171 will also play a part in future decisions about published enumerations.

DSEP recognized the need to develop ways to communicate with state stakeholders and the public about data protections that based on 2020 DAS methods. Our messaging will have to provide some simpler description of how the methods make changes to the attributes of the people in block counts, but still provide accurate and usable data.

DSEP noted that The National Conference of State Legislators (NCSL) will be expecting updates from Decennial based on 2018 testing outcomes in anticipation of 2020 releases of PL 94-171. It will be important to engage NCSL in discussions about 2020 DAS methods.

DSEP acknowledged that this and other details from ADRM's research were scheduled for discussion at the May 10, 2017 meeting of the 2020 Census Portfolio Management Governing Board (PMGB). DSEP postponed further discussion on this project and requests, pending any feedback from the presentation on this topic to the 2020 PMGB.

Post Meeting Notes:

DSEP revisited this topic at the beginning of the May 11, 2017 meeting.

Regarding issues of surrounding Voting Rights Acts Requirements, DSEP recognized that Decennial would need to talk to Justice if we were to alter any of the 6 constraints from PL 94-171 for 2020.

DSEP noted that the 2020 PMGB is supportive of the efforts of the 2020 DAS to optimize output noise infusion methods while publishing the most accurate data possible. There was unanimous support from 2020 PMGB for DRB's determination that the six data elements from PL 94-171 should be published as enumerated and form the base for the 2018 End-to-End testing research with the 2020 DAS.

DSEP agreed that the DRB should require that any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB.

Decision:

Request 1: DSEP approves publication of the six data elements from PL 94-171 as enumerated for the 2018 End-to-End test. Based on lessons learned, the use of these constraints for the PL 94-171 will be revisited for 2020.

Request 2: DSEP agreed that the DRB should require that any request for disclosure avoidance of proposed publications for the 2020 Census be routed to the 2020 DAS team before going to the DRB.

Record-level Re-identification Linkages for Evaluating the 2010 and 2020 Census Disclosure Avoidance Systems

Background:

The DAS team is attempting a database reconstruction using data from the 2010 PL94-171 and SF1 tabulations. The next step is to link those reconstructed microdata to commercial name and address files obtained in support of post-2010 research meant to represent the type of publically available file an attacker might potentially acquire. These files include Experian, InfoGroup, Melissa, Targus, TransUnion, and VSGL. This linkage involves the use of name and address data.

The final step is to compare the fully reconstructed microdata, including the commercially supplied names and address, to the name and address data on the 2010 Census Unedited File (CUF). Following accepted disclosure avoidance evaluation practices on re-identification, the 2020 DAS team would report to DRB and DSEP the putative re-identification rate (percentage of the records in the reconstructed microdata that could be linked to name and address information in the commercial files) and the proportion of putative re-identifications that were correct (proportion of reconstructed data records with putative re-identifications that were correctly linked to 2010 Census responses, including name and address).

Discussion:

DSEP recognized that the project proposal meets Data Linkage Policy requirements and involves sensitive but critical work that will allow the 2020 DAS subteam to understand the degree of risk of re-identification and database reconstruction with Census files.

DSEP noted that the subteam assembled for this research is composed of federal employees and one SSS individual.

Decision:

DSEP approved this project.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

DSEP Meeting Record

Topics: 2020 Decennial Record Linkage Test (Ned Porter, CSRM)

[REDACTED]

[REDACTED]

Meeting Date: May 11, 2017

<i>Position</i>	<i>Attending for Position</i>
Deputy Director (Chair)	Ron Jarmin
CAO	David Ziaya
CFO	Joanne Crane
AD, Decennial	Al Fontenot
AD, Demographic	Karen Battle
AD, Economic	Ron Jarmin
AD, Field	Joan Hill
AD, IT	Nitin Naik
AD, Research and Methodology	John Abowd
AD, 2020 Census	Al Fontenot
AD, Communications	Stephen Buckner
AD, Performance Improvement	Ted Johnson
Chief, PCO/ Chief Privacy Officer	Robin Bachman
Chief Demographer	Howard Hogan
Senior Advisor Designee from the Director's Office	<i>absent</i>
Chief Information Security Officer	Tim Ruland
Asst. Director, Research and Methodology	John Eltinge

Also Attending:	Simson Garfinkle, Tommy Wright, Eloise Parker, Ned Porter, Bill Winkler, Christa Jones, Letitia McKoy, Melissa Creech, Hampton Wilson, Ashley Landreth, Mike Castro, Janean Darden, Julie Atwell
-----------------	---

Administrative Notes:

At the beginning of the meeting, DSEP resumed their discussion and made a final decision on the topic: *Initial Request for DSEP Determination on Disclosure Avoidance for the 2018 End-to-End Test of the 2020 Census of Population and Housing*. The summary of that discussion and decision is in the May 10, 2017 meeting record.

2020 Decennial Record Linkage Test

Background:

Identifying duplicate records in the decennial census is critical to providing a more accurate count. One of the areas of research for improving the Decennial Matching methodology is improving the computer matching in the Duplicate Person Identification (DPI) process. This research will use the 2010 Census Unedited File (CUF) as well as data from Census Coverage Measurement (CCM). In addition, the research will determine if it is possible to increase the proportion of records receiving Personal Identification Keys (PIKs).

This research requires access to PIKs and complete name data on the files. This access will be limited to only five Census Bureau researchers as well as the Center for Statistical Research and Methodology's Data Steward. The data will be restricted to only authorized clusters.

Discussion:

DSEP acknowledged that research into deduplication methods is a routine and critical part of Census operations. DSEP further acknowledged that while this research project will use new technology and methods, it is fundamentally the same as research that happened in previous censuses.

Decision:

DSEP approved the project.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

DSEP Meeting Record

Topics:



Database Reconstruction Issue Mitigation (John Abowd, ADRM)

Meeting Date: February 15, 2018

<i>Position</i>	<i>Attending for Position</i>
COO (Chair)	Enrique Lamas
ADDC	Albert Fontenot
ADDP	Karen Battle
ADEP	Nick Orsini and Ron Jarmin
ADFO	Tim Olson
ADITCIO	Nitin Naik
ADRM	John Abowd
ADCOM	Stephen Buckner
CAO	David Ziaya
CFO	Joanne Crane
Asst. DRM	John Eltinge
Chief PCO/ Chief Privacy Officer	Robin Bachman
S.A. Director's Office	Douglas Clift
CISO	<i>Absent</i>
At-Large	Howard Hogan
At-Large	Frank Vitrano
Also Attending:	William Samples, David Waddington, Burton Reist, Victoria Velkoff, Robert Sienkiewicz, Jim Treat, Cynthia Hollingsworth, Clifford Jordan, Julia Naum, Jim Dinwiddie, Simson Garfinkel, Melissa Creech, Pat Cantwell, Byron Crenshaw, Hampton Wilson, Ashley Landreth, Mike Castro, Julie Atwell, Michael Snow

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Database Reconstruction Issue Mitigation

Background

The Census Bureau's Operating Committee (OPCOM), serving as the Enterprise Risk Review Board, elevated the enterprise risk of database reconstruction to an enterprise issue based on the results of a database reconstruction attack research effort the Census Bureau launched to understand that risk better. When an enterprise risk is elevated to an enterprise issue, the risk owner must implement an active mitigation plan to mitigate the risk. To that end, the Research and Methodology Directorate presented six recommendations to help manage the Census Bureau's publication strategy in ways that will protect its databases from reconstruction attacks.

NOTE: presenters and DSEP recognized that implementing several of the recommendations will require decisions on budget and staffing resources and that those decisions would need to be handled by other bodies at the Census Bureau. DSEP confined its discussion to establishing policy in response to the recommendations.

The following 6 recommendations were presented to DSEP:

1. **Suspension until September 30, 2019 of ad hoc releases of sub-state geography from any confidential source unless vetted differential privacy tools, or a DRB-approved noise-infusion alternative, have been used to produce the publication. This applies to all research projects whether they are external or internal. It does not apply to scheduled publications from sponsored survey clients for whom there is already an approved DRB protocol. Those clients should be put on notice for subsequent contracts. The complete list of approved exceptions, including sponsored survey products, is provided in 20180215b-External_Internal_Substate_Geography.xlsx. The suspension will be reviewed prior to September 30, 2019.**

NOTE: This suspension does not apply to state and national publications. It also does not apply to already scheduled publications from regular production activities. Program areas provided ADRM a list of those scheduled publications that should be exempted from the suspension. ADRM proposed ending those exemptions by September 30, 2019 even for those publications if they were not being produced using formally private systems by that point.

Discussion: DSEP recognized the need to modernize the Census Bureau's disclosure avoidance systems. DSEP acknowledged that by approving a list of exemptions they are agreeing to hold elevated levels of risk of database reconstruction associated with all of these data products. However, DSEP acknowledged the Census Bureau is obligated to provide the data the public needs for decision making and some of the release dates are required by law.

DSEP also acknowledged the need to set a target date for making these changes. While the ultimate goal is to make the publications of all of our programs formally private, that likely will not happen by September, 2019. However, in the meantime significantly improved noise infusion methods will be put in place to mitigate reconstruction risk.

DSEP members expressed concern that the list of already scheduled publications presented might be incomplete and asked for additional time for program areas to review the list and submit updates. DSEP agreed that the Center for Disclosure Avoidance Research (CDAR) should continue to accept submissions and finalize the list in advance of the next DSEP meeting. DSEP will formally approve the list at that point.

Decision: DSEP will finalize their approval of this recommendation at the March 15 DSEP meeting once the list of excepted publications has been finalized.

Action Items: Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23. The Chief of CDAR will redistribute the combined list to all contributors by February 28. CDAR will finalize the list of approved exceptions for distribution before DSEP's meeting on March 15.

2. **Suspension of all proposed tables in Summary File 1 and Summary File 2 for the 2020 Census at the block, block-group, tract, and county level except for the PL94-171 tables, as announced in Federal Register Notice 170824806–7806–01 (November 8, 2017, pp. 51805-6). To add a summary file table at any level of geography, racial/ethnic subpopulation other than OMB aggregate categories as specified in the 1997 standard (Federal Register October 30, 1997, pp. 58782-90), or group quarters type below the 2010 P42 seven categories, an affirmative case must be made for that table, use cases identified, and suitability for use standards developed. In addition, we recommend that the voting-age invariant in PL94-171 be removed, so that voting-age would be protected. DSEP will be asked to approve the SF1 and SF2 table specifications once they have cleared 2020 governance.**

NOTE: The PL94-17 tables from the 2018 End-to-End Census Test have been designed with a formally private system already and will be published, with the voting-age invariant, as planned.

Discussion: DSEP recognized that the SF1 and SF2 involved a very detailed set of tables that had been created to suit a wide set of data users. These tables were created, as a rule, to produce as much highly accurate data as possible within the existing disclosure avoidance framework. However, DSEP acknowledged that these data in many cases were accurate to a level that was not supported by the actual uses of those data, and such an approach is simply untenable in a formally private system.

DSEP acknowledged a fundamental need to take stock of what data the Census Bureau is required to publish, both by statute and the needs of our data users, and at what level of accuracy. This is not an activity that should be done by our Disclosure Review Board. Program areas have to make the case of what the data will be used for, and the actual minimum level of accuracy needed for those uses, so that CDAR and the DRB can build the system to allocate the privacy-loss budget according to those use cases.

A redesign of SF1 and SF2 based on formally articulated use cases will take a tremendous amount of effort but cannot be done in a vacuum. Program areas will have to reach out to data-user communities on developing the use cases for the needed data accuracy and levels of geography.

NOTE: DSEP discussed but tabled until later any decision on changing the voting-age invariant for the PL94-171 table produced as part of the 2020 Census.

Decision: DSEP approved this recommendation. For the 2020 Census, SF1 and SF2 will be rebuilt based on use cases.

Action Items: DCMD, POP, and ADDC divisions will work with the relevant program management governing board (PMGB) to establish a plan to execute this redesign.

3. Immediate review of all sub-state geography scheduled publications from the American Community Survey (ACS) to determine which ones can be delayed until there is a formally private publishing system for ACS.

Discussion: DSEP acknowledged that many of the ACS tables are already in production and that production needs to move forward. DSEP acknowledged that there are likely no publications currently suitable for delay, however they emphasized that ACSO needs to ensure that all exceptions are added to the list.

Decision: DSEP approved this recommendation.

Action Items: ACSO will verify that they have included all of the necessary publications on the list of exempted data releases.

4. Consideration of postponing ACS PUMS releases indefinitely.

NOTE: DSEP recognized that all of the publication systems and methods for the Census of Island Areas are identical to the ACS. DSEP emphasized that any changes made to the ACS should also reflect consideration of the needs of the Island Areas.

Discussion: DSEP acknowledged that while the threat of database reconstruction and reidentification attacks applies to all of the Census Bureau's data products, should the ACS data be subject to a reidentification attack, from a public perception standpoint, our continued publication of the ACS PUMS files would appear to be an egregious mistake.

However, DSEP also acknowledged that the ACS PUMS is a heavily used dataset for research and recognized that discontinuing this publication could generate a great deal of traffic for the FSRDCs. DSEP acknowledged that, before the Census Bureau restricts use the ACS PUMS to the FSRDCs, it needs to verify that they can handle the increased workload. Additionally, at present there are no FSRDCs that are readily accessible from the Island Areas.

DSEP recognized that immediate suspension of the ACS PUMS would cause a great deal of concern among data users and others. DSEP discussed the need to work on messaging around

any suspension and to brief the Department of Commerce before the Census Bureau implements the suspension.

Decision: DSEP deferred for one month any decisions to suspend release of the ACS PUMS pending further consideration of the ability of the FSRDC network to support increased demand, the impact on the data needs of the Island Areas, and development of a messaging plan.

Action Items: ADRM will prepare an assessment of the potential increased demand on the FSRDC network, and Decennial will prepare an assessment of the impact of suspending this publication on the Island Areas. ADCOM will work on a messaging plan.

5. Mandate for the 2022 Economic Censuses to use formally private publication systems for all tables.

Discussion: DSEP recognized that it is too late to begin creating a formally private system for data releases from the 2017 Economic Census. DSEP additionally discussed how modernizing disclosure avoidance systems will involve much more than just budgeting extra funds. It also will require having the adequate number of people with the right skills to do the work.

DSEP recognized that program areas will have to involve their PMGB in setting resources, budgets, and timelines and that it should be feasible to put formally private systems in place in time for the 2022 Economic Census.

Decision: DSEP approved this recommendation. The Census Bureau will move forward with designing and implementing formally private systems for the 2022 Economic Census.

6. Mandate to the Demographics Directorate to begin negotiations with survey clients for increased use of restricted-access microdata protocols and formally private table publication systems.

POST MEETING NOTE: a member in attendance recommended that there should also be outreach to reimbursable clients for the Economic Directorate.

Discussion: DSEP recognized the need to begin discussions with sponsors of Census Bureau surveys but determined that the Census Bureau should have a communications plan in place before mandating that the Demographic Directorate speak to sponsors.

Decision: DSEP will reconsider in one month whether to mandate conversations with survey and report sponsors.

Consolidated Action items:

- Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23.
- The Chief of CDAR will redistribute the combined list to all contributors by February 28.
- DCMD, POP, and the ADDC will work with the relevant PMGBs to establish a plan to execute the redesign of SF1 and SF2 based on use cases.
- ACSO will work to determine that all ACS data releases in production are listed on the spreadsheet of exceptions to the suspension.
- ADRM will prepare an assessment of the potential increased demand on the FSRDC network from suspension of the ACS PUMS.
- ADCOM will work on a messaging plan related to the suspension of the ACS PUMS.
- Decennial will prepare an assessment of the impact of suspending publication of the ACS PUMS on the Island Areas.

Staring Down the Database Reconstruction Theorem

John M. Abowd
Chief Scientist and Associate Director for Research and Methodology
U.S. Census Bureau
American Association for the Advancement of Science
Annual Meeting Saturday, February 16, 2019 3:30-5:00



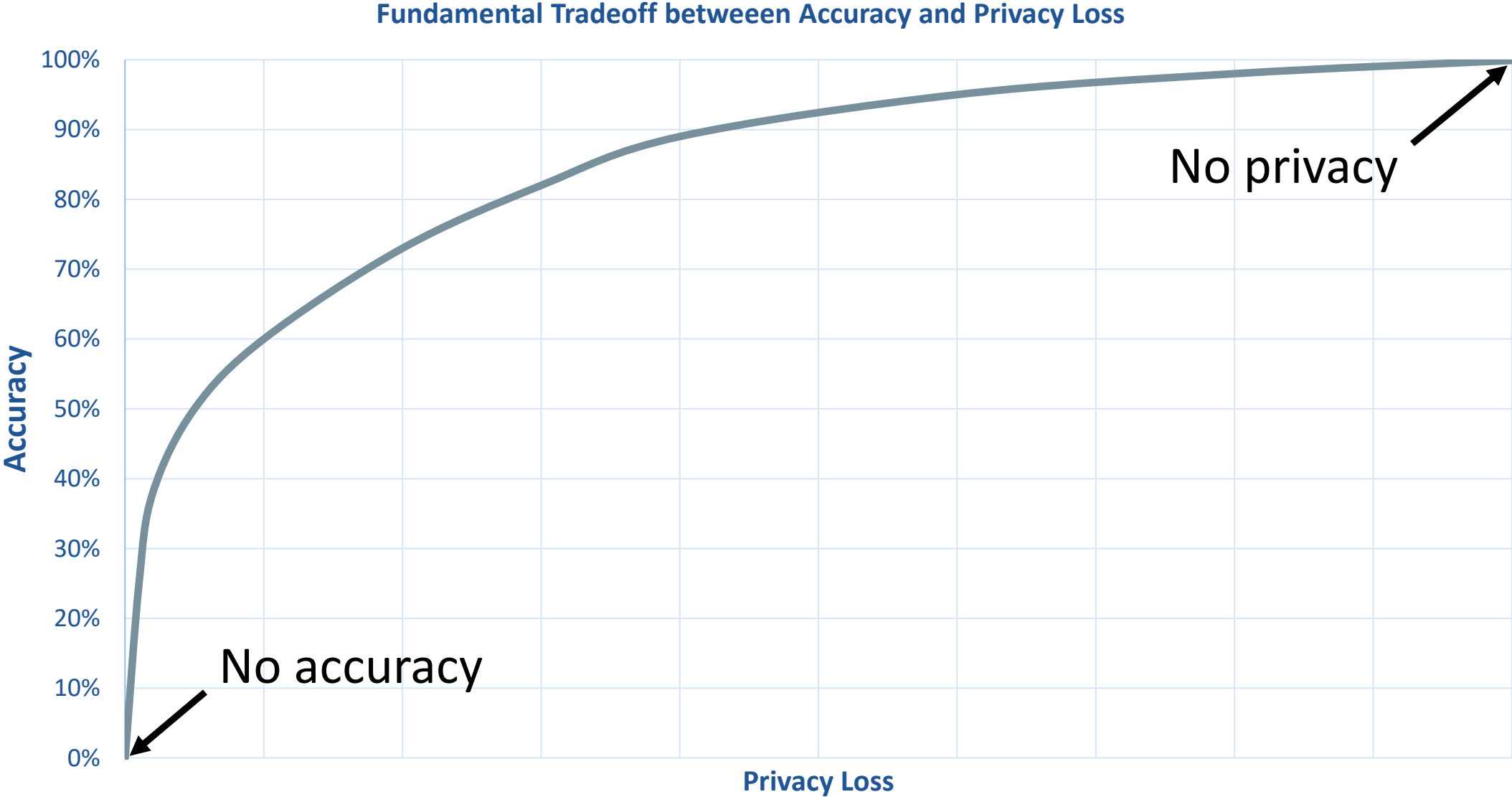
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

The views expressed in this talk are my own
and not those of the U.S. Census Bureau.

The challenges of a census:

1. collect all of the data necessary to underpin our democracy;
2. protect the privacy of individual data to ensure trust and prevent abuse.

- Too many statistics
- Noise infusion is necessary
- Transparency about methods helps rather than harms



Good science and privacy protection are partners

Start

Base Map

Selection

Results

Distance/Direction Analysis

Work to Home

Display Settings

Labor Market Segment

Filter

Year

Map Controls

Color Key

Thermal Overlay

Point Overlay

Selection Outline

Identify

Clear Overlays

Zoom to Selection

Animate Overlays

Report/Map Outputs

Detailed Report

Export Geography

Print Chart/Map

Legends

Change Settings

Job Counts by Distance/Direction in 2015
All Workers

View as **Radar Chart**

Jobs by Distance - Work Census Block to Home Census Block

	2015	
	Count	Share
Total Primary Jobs	12,260	100.0%
Less than 10 miles	5,949	48.5%
10 to 24 miles	2,987	24.4%
25 to 50 miles	1,451	11.8%
Greater than 50 miles	1,873	15.3%

Census Bureau

Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Privacy Policy | 2010 Census | Data Tools | Information Quality | Product Catalog | Contact Us | Home
Source: U.S.Census Bureau, Center for Economic Studies | e-mail: CES.OnTheMap.Feedback@census.gov

6

What we did

- Database reconstruction for all 308,745,538 people in 2010 Census
- Link reconstructed records to commercial databases: acquire PII
- Successful linkage to commercial data: putative re-identification
- Compare putative re-identifications to confidential data
- Successful linkage to confidential data: confirmed re-identification
- Harm: attacker can learn self-response race and ethnicity

What we found

- Census block correctly reconstructed in all 6,207,027 inhabited blocks
- Block, sex, age, race, ethnicity reconstructed
 - Exactly: 46% of population (142 million of 308,745,538)
 - Allowing age +/- one year: 71% of population (219 million of 308,745,538)
- Block, sex, age linked to commercial data to acquire PII
 - Putative re-identifications: 45% of population (138 million of 308,745,538)
- Name, block, sex, age, race, ethnicity compared to confidential data
 - Confirmed re-identifications: 38% of putative (52 million; 17% of population)
- For the confirmed re-identifications, race and ethnicity are learned exactly, not statistically

We fixed this for the 2020 Census by implementing differential privacy

Acknowledgments

- The Census Bureau's 2020 Disclosure Avoidance System incorporates work by Daniel Kifer (Scientific Lead), Simson Garfinkel (Senior Scientist for Confidentiality and Data Access), Rob Sienkiewicz (ACC Disclosure Avoidance, Center for Enterprise Dissemination), Tamara Adams, Robert Ashmead, Michael Bentley, Stephen Clark, Craig Corl, Aref Dajani, Nathan Goldschlag, Michael Hay, Cynthia Hollingsworth, Michael Ikeda, Philip Leclerc, Ashwin Machanavajjhala, Christian Martindale, Gerome Miklau, Brett Moran, Edward Porter, Sarah Powazek, Anne Ross, Ian Schmutte, William Sexton, Lars Vilhuber, Cecil Washington, and Pavel Zhuralev

Thank you.

John.Maron.Abowd@census.gov

More Background on the 2020 Census Disclosure Avoidance System

- September 14, 2017 CSAC (overall design)
<https://www2.census.gov/cac/sac/meetings/2017-09/garfinkel-modernizing-disclosure-avoidance.pdf?#>
- August, 2018 KDD'18 (top-down v. block-by-block)
<https://digitalcommons.ilr.cornell.edu/ldi/49/>
- October, 2018 WPES (implementation issues)
<https://arxiv.org/abs/1809.02201>
- October, 2018 [ACMQueue](#) (understanding database reconstruction)
<https://digitalcommons.ilr.cornell.edu/ldi/50/> or
<https://queue.acm.org/detail.cfm?id=3295691>
- December 6, 2010 CSAC (detailed discussion of algorithms and choices)
<https://www2.census.gov/cac/sac/meetings/2018-12/abowd-disclosure-avoidance.pdf?#>

Selected References

- Dinur, Irit and Kobbi Nissim. 2003. Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '03). ACM, New York, NY, USA, 202-210. DOI: 10.1145/773153.773173.
- Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. In Halevi, S. & Rabin, T. (Eds.) Calibrating Noise to Sensitivity in Private Data Analysis Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings, Springer Berlin Heidelberg, 265-284, DOI: 10.1007/11681878_14.
- Fellegi, Ivan P. 1972. On the Question of Statistical Confidentiality. Journal of the American Statistical Association, Vol. 67, No. 337 (March):7-18, stable URL <http://www.jstor.org/stable/2284695>.
- Ganda, Srivatsava, Shiva Kasiviswanathan and Adam Smith. 2008. Composition Attacks and Auxiliary Information in Data Privacy. In Knowledge, Discovery and Datamining, Las Vegas, NV, doi:10.1145/1401890.1401926.
- Machanavajjhala, Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. Privacy: Theory Meets Practice on the Map, International Conference on Data Engineering (ICDE) 2008: 277-286, doi:10.1109/ICDE.2008.4497436.
- McKenna, Laura. 2018. Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing, Working Papers 18-47, Center for Economic Studies, U.S. Census Bureau, Handle: RePEc:cen:wpaper:18-47.
- Ramachandran, Aditi, Lisa Singh, Edward Porter, and Frank Nagle. 2012. Exploring Re-Identification Risks in Public Domains, Tenth Annual International Conference on Privacy, Security and Trust, IEEE, doi:10.1109/PST.2012.6297917.
- U.S. Census Bureau. 2019. LEHD Origin-Destination Employment Statistics (2002-2015) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on February 15, 2019 at <https://onthemap.ces.census.gov>.

[Slide 1] [Before I start, I want to remind members of the audience that, while I am appearing in my official capacity as the Chief Scientist of the U.S. Census Bureau, I am presenting a summary of research findings. The views expressed in this talk are my own, not those of the Census Bureau.]

Staring Down the Database Reconstruction Theorem

[Slide 2] The 2020 Census will be the safest and best-protected ever. This is not nearly as easy as it sounds.

Throughout much of the history of the decennial census, our country has struggled with two challenges:

- 1) collect all of the data necessary to underpin our democracy;
- 2) protect the privacy of individual data to ensure trust and prevent abuse.

The first obligation derives directly from the Constitution, of course. As for the privacy requirement, Section 9 of the Census Act (Title 13 of the U.S. Code) prohibits making “any publication whereby the data furnished by any particular establishment or individual under this title can be identified.” In fact, the Census Bureau is about the only organization operating under a blanket U.S. legal requirement never to release data that can be tied back to individuals or companies no matter what.

The Census Bureau has always been committed to meeting both of its obligations; that is, providing population statistics needed by decision-makers, scholars, and businesses while also protecting the privacy of census participants.

A paper by Laura McKenna (2018), who supervised the confidentiality protection systems used by the Census Bureau for more than 15 years, catalogued the public information about the technical systems used for protection of publications from decennial censuses since 1970.

As McKenna noted, beginning with the 1990 Census, the primary confidentiality protection method employed was household-level swapping of geographic identifiers—moving an entire household from one location to another—prior to tabulating the data. The goal was to introduce uncertainty about whether households allegedly re-identified from the published data were correct.

Essentially the same methods were used for the 2000 and 2010 Censuses but with refinements that recognized the changing external environment.

The discipline of statistics has evolved over the last century. So too has the widespread availability of data. With each new development, the Census Bureau must ask how the current state of affairs will affect the production of the statistical products that it releases to the public so as to be both useful and privacy-preserving.

Sixteen years ago, two computer scientists, Irit Dinur and Kobbi Nissim (2003), wrote a seminal article proving a “database reconstruction theorem,” which is also known as the “fundamental law of information recovery.”

Three years later, Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith (2006) provided a mathematical foundation for what we now call “differential privacy.” In short, they explained how to quantify the limits on the accuracy of answers to queries based on the confidential data and the privacy-loss to the entities in those data, when the queries are answered publicly. More importantly, they provided a technique for enhancing privacy that goes far beyond the swapping approach that many statisticians have been using for years.

[Slide 3] The full implications of database reconstruction were not understood in 2003, but over the next several years a scientific consensus emerged in the data privacy community that:

- **Too many statistics**, published too accurately, expose the confidential database with near certainty (Dinur and Nissim 2003).
- **A necessary condition for controlling privacy loss** against informed attackers is to add noise to every statistic, calibrated to control the worst-case disclosure risk, which is now called a privacy-loss budget (Dwork, McSherry, Nissim and Smith 2006; Ganta, Kasiviswanathan, and Smith 2008).
- **Transparency about methods helps rather than harms**, Kerckhoff’s principle, applied to data privacy, says that the protections should be provable and secure even when every aspect of the algorithm and all of its parameters are public. Only the actual random number sequence must be kept secret (Dwork, McSherry, Nissim, and Smith 2006).

If you curate confidential data, then you can use those data for two competing goals:

- You can publicly and precisely answer statistical queries about the data.
- You can preserve and protect the privacy of those whose information is in the data.

You can do some of both.

[Slide 4] But if you do all of one, you can't do any of the other.

Period.

This trade-off is one of the hardest lessons to learn in modern information science. It is a lesson about data generally, not about counting people. And it is a mathematical theorem, not an opinion or implementation detail.

[Slide 5] This transformation in the fields of statistics and computer science is truly mind-blowing. It's at the heart of the science that we're here to celebrate. Cryptographers usually study the safety of methods for encrypting information about private data. Now their insights show us safe ways to publish information from private data. The cryptographic approach shows that some new methods can provably protect privacy, and some old methods provably do not. But the safe methods only work if we accept the inherent limitations on the accuracy of those publications that the cryptographers have highlighted.

Specifically, technical advances revealed a new vulnerability, allowing people to reconstruct data from tables that were previously assumed to be privacy-preserving, given the available computing resources. But other technical advances have also enabled a new form of privacy protection that is not only more sophisticated but also mathematically grounded in a way that allows statisticians to fully understand the limits of what they can make available and what kind of privacy they can provably offer. This dual breakthrough is transforming how we protect data today.

Good science and real privacy protection turn out to be partners, not competitors, in the efforts to modernize the methods data analysts use. For this reason, we have seen many companies, like Google, Microsoft, and Apple, turn to differential privacy to secure data and make guarantees about the privacy of

statistical tables. But it was actually the Census Bureau who first recognized the power of this method at scale.

[Slide 6] In 2008, the Census Bureau implemented an early version of differential privacy on data that display the commuting patterns of people based on where they live and work (Machanavajjhala et al. 2008; U.S. Census Bureau 2019).

Working with statisticians and computer scientists, we have collectively advanced the state of differential privacy such that we are going to implement it at scale as part of the 2020 Census. While I will talk about what that looks like in more detail tomorrow at 8:00AM, today I want to explain why we absolutely must implement differential privacy in order to protect the privacy of those participating in the census.

Starting in 1972, researchers began highlighting how it was possible to combine statistical tables and use differencing techniques to identify which census respondents provided the associated data (Fellegi 1972). As the market for detailed data grew and evolved, researchers also began highlighting how combining commercial data with census tables could introduce new vulnerabilities. While external users could not provably know whether or not their reconstructions were accurate, the Census Bureau recognized that it was critical to know the potential vulnerability of census data.

We acted proactively, as the Census Bureau has done for many decades. We designed our own internal research program to assess the current state of this vulnerability without waiting for a specific external threat. I'm now going to explain what we found.

[Slide 7] Here are the steps we followed:

- Using only published contingency tables (summary statistics), we applied the database reconstruction theorem to construct record-level images for all 308,745,538 persons enumerated in the 2010 Census. A record-level image is a row in the reconstructed database with the same variables that were used in publications from the confidential database. There is no traditional PII (personally identifiable information) on these reconstructed records.

- Using only the information in the reconstructed data records, we linked those records to commercial databases to acquire name and address information. This information would have been available to an external attacker, circa 2010.
- When the record linkage operation is successful, the PII from the commercial data are attached to the reconstructed census record. We call the reconstructed record, now laden with PII, “putatively re-identified,” which means that an attacker might think that the attack was successful.
- We then compared the putatively re-identified census records to the real confidential census records. When this comparison matched on all variables, including the PII and those variables not available in the commercial data, we called this a “confirmed re-identification.”
- The harm from such re-identifications, in the 2010 Census, is that the attacker learns the self-reported race and ethnicity on the confidential census record. Those data are not available in identifiable form to any commercial or governmental agency except the Census Bureau.

[Slide 8] Here are the basic results:

- In the reconstructed data, certain variables are always correctly reconstructed—meaning that the value in the reconstructed variable always matches its value in the confidential data. The census block, where the person lived on April 1, 2010, is always correctly reconstructed. This is true for every one of the 6,207,027 inhabited blocks in the 2010 Census.
- All the variables we studied: block, sex, age in years, race, and ethnicity are exactly correct in the reconstructed records for 46% of the population (142 million of 308,745,538 persons)—meaning that the reconstructed record exactly matches the confidential record on the value of all five variables. This result is salient because in the confidential data, more than 50% of the records are unique in the population—the only instance of this combination of values observed in the census (the exact percentage is confidential). If we allow the age to vary by plus or minus one year, then the number of reconstructed records that match the confidential data on these five variable rises to 71% (219 million of 308,745,538 persons).
- When we use the reconstructed block, sex and age to link each reconstructed record to the records harvested from commercial data

acquired at the time of the 2010 Census, we putatively re-identify 45% of the total population (138 million of 308,745,538 persons). That means that we were able to attach a unique name and address to 45% of the reconstructed records from the 2010 Census. The match is exact for block and sex. Age is allowed to vary by plus or minus one year.

- When we compared the unique name, block, sex, age, race, and ethnicity on the putative re-identifications to the same variables on the 2010 Census confidential data, we confirmed 38% of these matches (52 million of 308,745,538 persons, or 17% of the total population).

The putative re-identifications probably have a recall rate (or sensitivity) of at least 45%. Neither the attacker nor the Census Bureau have PII on all 308,745,538 persons enumerated in the 2010 Census, so the correct recall rate denominator is certainly less than the total population.

The precision of the record linkage is 38%, which means that the attacker would be correct between one-quarter and one-half of the time.

And both of these estimates (45% putatively re-identified; 38% of which are correct) are really lower bounds for other reasons: our experiments didn't use all of the information that the Census Bureau published from the 2010 Census. For example, we didn't use any information on household composition, which means that potential harm from discovering other features of households, like same-sex unions and adoptions, is still unquantified. We also made no use of the 2010 Public-Use Microdata Sample.

To further put these results in context, the last time the Census Bureau released results for a re-identification study, which did not use database reconstruction (Ramachandran et al. 2012), the putative re-identification rate was 0.017% (389 persons of 2,251,571) and the confirmation rate was 22% (87 of 389).

[Slide 9] All of us—the entire scientific community—have an obligation to examine the methods we use in light of the cryptographic critique of the privacy protections those methods offer. We must also recognize that these developments are sobering to everyone.

This is not just a challenge for statistical agencies or Internet giants, although those institutions have been in the vanguard of this movement.

It's a challenge for Internet commerce, because recommendation systems expose private data.

It's a challenge for bioinformatics, because summaries of genomes expose private data.

It's a challenge for commercial lenders, because benchmark risk assessments expose private data.

It's a challenge for nonprofit survey organizations, because their research reports expose private data.

Regardless of what anyone says, people want to be assured that their data are private. They want to know that we can't use statistical magic to re-identify information that they thought was private. They want to know that statistical tables can't come back to haunt them.

That's why I'm so grateful that the data we are showing today aren't the end of the story. They simply show that we cannot accept the status quo. We cannot presume that what worked a decade ago will work again in 2020. We have to innovate. And that's what we are doing.

In 2016, the Census Bureau acknowledged that database reconstruction was a vulnerability of the methods traditionally used to protect confidentiality in decennial census publications.

What we showed today is that we have a clear understanding of how it's possible to reconstruct 2010 Census data for block, sex, age, race and ethnicity. But this understanding isn't in vain. This understanding gave us the information we needed to develop techniques to make sure this isn't possible in 2020.

We are going into the 2020 Census confident that we can protect the privacy of all who participate. We have to make some important decisions about what statistics should be made available and how to weigh public data interests with our commitment to keep individual data private from reconstruction. But we know where the vulnerabilities are and we have the tools to make certain that what I showed today can't happen in the future.

The publications of the 2020 Census will be protected by differential privacy because it's imperative above all else that we ensure the trust of the American people.

The exact algorithms, and all parameters, will also be publicly released well in advance of the tables because it is imperative that we be accountable to the scientific community and the public at large.

[Slide 10] Statistics has evolved significantly over the last century. I'm honored to be a part of a statistical agency with a long tradition of implementing cutting-edge knowledge on the behalf of the American people. And I'm deeply grateful to the amazing team at the Census Bureau for identifying the challenges we face and ensuring that we can meet those challenges.

I promise the American people that they will have the privacy they deserve.

For those who would like to know more about how we are implementing differential privacy in the 2020 Census, please join me tomorrow at 8:00 AM where I will present our methods in more detail.

References

- Dinur, Irit and Kobbi Nissim. 2003. Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '03). ACM, New York, NY, USA, 202-210. DOI: 10.1145/773153.773173.
- Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. In Halevi, S. & Rabin, T. (Eds.) Calibrating Noise to Sensitivity in Private Data Analysis Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings, Springer Berlin Heidelberg, 265-284, DOI: 10.1007/11681878_14.
- Fellegi, Ivan P. 1972. On the Question of Statistical Confidentiality. Journal of the American Statistical Association, Vol. 67, No. 337 (March):7-18, stable URL <http://www.jstor.org/stable/2284695>.
- Ganda, Srivatsava, Shiva Kasiviswanathan and Adam Smith. 2008. Composition Attacks and Auxiliary Information in Data Privacy. In Knowledge, Discovery and Datamining, Las Vegas, NV, doi:10.1145/1401890.1401926.
- Machanavajjhala, Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. Privacy: Theory Meets Practice on the Map, International Conference on Data Engineering (ICDE) 2008: 277-286, doi:10.1109/ICDE.2008.4497436.
- McKenna, Laura. 2018. Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing, Working Papers 18-47, Center for Economic Studies, U.S. Census Bureau, Handle: RePEc:cen:wpaper:18-47.

Abowd, AAAS presentation Saturday, February 16, 2019, 3:30-5:00

FINAL Page 9

Ramachandran, Aditi, Lisa Singh, Edward Porter, and Frank Nagle. 2012. Exploring Re-Identification Risks in Public Domains, Tenth Annual International Conference on Privacy, Security and Trust, IEEE, doi:10.1109/PST.2012.6297917.

U.S. Census Bureau. 2019. LEHD Origin-Destination Employment Statistics (2002-2015) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on February 15, 2019 at <https://onthemap.ces.census.gov>.

**IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
EASTERN DIVISION**

STATE OF ALABAMA, *et al.*,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF
COMMERCE, *et al.*,

Defendants.

Case No. 3:21-CV-211-RAH-ECM-KCN

DECLARATION OF MICHAEL THIEME

I, Michael Thieme, make the following Declaration pursuant to 28 U.S.C. § 1746, and state that under penalty of perjury the following is true and correct to the best of my knowledge and belief:

1. I am the Assistant Director for Decennial Census Programs, Systems, and Contracts at the U.S. Census Bureau. I have occupied this position since November 2017. The 2020 Census is my third Decennial Census. For the 2010 Census, I was the Chief of the Decennial Systems and Contracts Management Office, providing the primary technology and contract management support for that census. For the 2000 Census, I was the Special Assistant to the Assistant Director for Field Operations working at the national level directing field data collection. In my current role as Assistant Director I am responsible for three Census Bureau divisions: the Decennial Information Technology Division, the Geography Division, and the Decennial Contracts Execution Office. With over 2,000 employees and contractors, these divisions provide all the information technology, geography, and contract management support for the 2020 Census. I am knowledgeable about the progress of the 2020 Census in general and the processing of census data in particular.

2. I am making this Declaration in support of Defendants' Opposition to Alabama's preliminary-injunction motion. All statements in this Declaration are based on my personal knowledge or knowledge obtained in the course of my official duties. In this declaration I:

- Provide background about the progress of the 2020 Census and delays;
- Stress the Census Bureau's commitment to producing high quality, usable, data products from the 2020 Census; and
- Provide background on how the Census Bureau processes data for the 2020 Census and why we are unable to produce redistricting data before the statutory deadline.

Background on the 2020 Census

3. The Census Bureau goes to extraordinary lengths to count everyone living in the country once, only once, and in the right place. The Census Bureau's goal in conducting the decennial census is to count everyone living in the United States, including the 50 states, the District of Columbia, and the territories of Puerto Rico, American Samoa, Commonwealth of the Northern Mariana Islands, Guam, and U.S. Virgin Islands. To that end, we expend significant funds, efforts, and resources in capturing an accurate enumeration of the population, including those who are hard to count.

4. The planning, research, design, development, and execution of a decennial census is a massive undertaking to count over 330 million people across 3.8 million square miles. The 2020 decennial census consisted of 35 operations using 52 separate systems. We monitored and managed the status and progress of the 2020 Census in large part using a master schedule, which has over 27,000 separate lines of census activities. Thousands of staff at Census Bureau headquarters and across the country supported the development and execution of the 2020 census operational design, systems, and procedures. In addition, the 2020 Census required the hiring and management of hundreds of thousands of field staff across the country to manage operations and collect data in support of the decennial census.

5. The complexity and inter-related nature of census operations is echoed in the budget for the 2020 Census. The overall budget estimate for the 2020 Census—covering fiscal years 2012 to 2023—was \$15.6 billion. The Government Accountability Office (GAO) determined that, as of January 2020, this estimate substantially or fully met GAO's standards and best practices for a reliable cost estimate in terms of credibility, accuracy, completeness, and documentation quality. It is rare for civilian agencies to be so designated, and we are proud that the Census Bureau has achieved this status. As of this writing, the Census Bureau has been appropriated in aggregate just over \$14.2 billion to use for the 2020 Census, covering fiscal years 2012 through 2021.

6. The operational design of the 2020 Census was subjected to repeated and rigorous testing. Given the immense effort required to conduct the census, the importance of the results, and the decade of work by thousands of people that goes into planning and conducting the decennial census, the Census Bureau expends a significant amount of effort to evaluate its planning and design to ensure that its operations will be effective in coming as close as possible to a complete count of everyone living in the United States. Design and testing of the 2020 Census was an iterative process: after each test, we revised our plans and assumptions as necessary.

7. The [2020 Census Operational Plan](#) explains the overall operations of the 2020 Census, including the integration of numerous sub-operations. Further details on most of these sub-operations can be found [on our website](#). A partial list of the major operations for which we have posted detailed operations plans includes:

- a. Local Update of Census Addresses
- b. Address Canvassing
- c. Geographic Delineations
- d. Field Infrastructure and Logistics
- e. Forms Printing and Distribution
- f. Integrated Communications Plan
- g. Count Review
- h. Intended Administrative Data Use
- i. Internet Self-Response
- j. Counting Federally Affiliated Americans Overseas
- k. Non-ID Processing
- l. Update Enumerate
- m. Update Leave
- n. Nonresponse Followup (NRFU)
- o. Response Processing

- p. Formal Privacy Methods
- q. Redistricting Data Program
- r. Post Enumeration Survey (PES)
- s. Count Question Resolution
- t. Data Products and Dissemination
- u. Evaluations and Experiments
- v. Archiving

Census Step 1: Locating Every Household in the United States

8. The first operational step in conducting the 2020 Census was to create a Master Address File (MAF) to represent the universe of addresses and locations to be counted in the 2020 Census. A national repository of geographic data—including addresses, address point locations, streets, boundaries, and imagery—is stored within the Census Bureau’s Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) System, which provides the foundation for the Census Bureau’s data collection, tabulation, and dissemination activities. It is used to generate the universe of addresses that will be included in a decennial census. Those addresses are then invited to respond, typically through an invitation in the mail. The MAF/TIGER System provides the address and geographic base used by our operational control systems to control responses as they are returned to the Census Bureau. The MAF/TIGER data are used to ensure that each person is tabulated to the correct geographic location as the final 2020 Census population and housing counts are prepared.

9. The Census Bureau continually updated this address list in preparation for the 2020 Census. For the third decade, as mandated by the Census Address List Improvement Act of 1994, the Census Bureau implemented the Local Update of Census Addresses (LUCA) Program to provide tribal, state, and local governments an opportunity to review and update the Census Bureau’s address list for their respective jurisdictions. Between

September 2015 and June 2017, the Census Bureau conducted a 100 percent in-office review of every census block in the nation (11,155,486 blocks).¹ During the in-office review, clerical staff had access to satellite and aerial imagery from federal, state, and local sources, and to publicly available street-level images through Google Street View and Bing StreetSide, which provided the ability to see the fronts of structures, as if standing on the sidewalk.

10. A field operation called In-Field Address Canvassing occurred between August 2019 and October 2019 for approximately 50 million addresses that were not verified in the in-office review. Address Canvassing fieldwork validated roughly 88% of these addresses and the remainder were removed from the universe because the Address Canvassing fieldwork verified that they did not exist, were duplicates, or were non-residential addresses. Some new addresses identified during fieldwork matched addresses already in the MAF as a result of contemporaneous in-office update processes. Other new addresses were added to the MAF.

11. The Census Bureau believes that the Census Bureau's MAF/TIGER System is the most complete and accurate address listing in census history.

Census Step 2: Encouraging Self-Response Throughout the 2020 Census

12. In order to encourage everyone in the United States to self-respond, the Census Bureau designed, tested, and implemented a \$700 million Integrated Communications Program. This included a massive multimedia campaign designed to engage

¹ Statistical geographies establish the geographic areas at which the Census Bureau produces statistics. Census blocks are the smallest geographic areas for which we collect and tabulate data. Census blocks are formed by streets, railroads, bodies of water, and legal boundaries (there are approximately 8 million Census Blocks). Census blocks are aggregated to form block groups, and block groups are aggregated to form census tracts. Census tracts optimally represent about 1,600 housing units and 4,000 people. These statistical geographies nest within governmental unit boundaries, such as municipalities and counties.

stakeholders and partners, and to communicate the importance of the census through paid advertising, public relations, social media content, and the new web site. This was the first census where we made a significant investment in digital advertising, targeting online sites including Facebook, Instagram, paid search engines, display ads, and programmatic advertising.

13. The Census Bureau adapted its outreach strategies in response to delayed census operations due to COVID-19, increasing advertising and outreach to specific areas of the country with lower response rates. We quickly adjusted our messaging, pivoting from our original campaign to encourage people to respond online from the safety of their own homes. The use of micro-targeting allowed the Census Bureau to tailor its messaging, including directing appropriate messages to hard-to-reach communities and those who distrust government, both of which have been traditionally undercounted.

14. The Census Bureau's communications program also relied heavily on partnerships, including with organizations in the State of Alabama. There are two prongs to the Partnership Program, the National Partnership Program that works from Census Bureau headquarters mobilizing national organizations, and the Community Partnership and Engagement Program, that works through the regions at the local level to reach organizations that directly touch their communities. Census partners include national organizations like the National Urban League (NUL), the Mexican American Legal Defense Fund (MALDEF), the National Association of Latino Elected Officials (NALEO), the National Association for the Advancement of Colored People (NAACP), and the U.S. Chambers of Commerce. Major corporations also become census partners. At the local level, partners can be churches, synagogues and mosques, legal aid clinics, grocery stores, universities, colleges, and schools.

Census Step 3: Self-Response

15. The design of the 2020 Census depended on self-response from the American public. In an effort to ensure the most efficient process to enumerate households, the

Census Bureau assigned every block in the United States to one specific type of enumeration area (TEA). The TEA reflects the methodology used to enumerate the households within the block. There were two TEAs where self-response was the primary enumeration methodology: TEA 1 (Self-Response) and TEA 6 (Update Leave). Regardless of enumeration methodology, everyone in the country was able to participate in the census online, by mail, or by phone.

16. TEA 1 used a stratified self-response contact strategy to inform and invite the public to respond to the census, and to remind nonresponding housing units to respond. In total, six mailings including the initial Invitation, reminders, and, if we did not receive a response by the third mailing, questionnaires were to be delivered on a flow basis unless a household responded.

17. Update Leave (TEA 6) was conducted in areas where the majority of the housing units did not have mail delivery to the physical location of the housing unit, or the mail delivery information for the housing unit could not be verified. The purpose of Update Leave was to update the address list and feature data, and to leave a 2020 Census Internet Choice package at every housing unit. The major difference from TEA 1 is that a Census Bureau employee, rather than a postal carrier, delivers the 2020 Census invitation to respond, along with a paper questionnaire. As with other housing units, those in TEA 6 had the option to respond online, by mail, or by phone.

18. Self-response began in March 2020 and was open until October 15, 2020. We are proud to have secured a self-response rate of 67%, higher than the 2010 self-response rate of 66.5%.

Census Step 4: Nonresponse Followup (NRFU) and Quality Control

19. After giving everyone an opportunity to self-respond to the census, census field staff (known as enumerators), attempted to contact nonresponding addresses to determine whether each address was vacant, occupied, or did not exist, and when occupied, to collect census response data. Multiple contact attempts to nonresponding addresses

were done to determine the housing unit status and to collect decennial census response data. This was the Nonresponse Followup operation, or NRFU. Enumerators conducted the NRFU operation using iPhones equipped with “optimization” software that assigned cases based on the enumerator’s availability and to increase efficiency of the operation.

20. In addition to the NRFU operation, the Census Bureau conducted several operations to collect information for individuals who do not live in housing units. The Group Quarters Enumeration collects response information for individuals living in group housing situations, such as college dormitories, prisons, or long term care facilities. The Enumeration at Transitory Locations (ETL) operation collects response information for individuals living at campgrounds and marinas.

21. Cases in the NRFU workload are subject to six contact attempts. The first contact attempt is primarily an in-person attempt. Each contact attempt in the 2020 Census NRFU was either a telephone or an in-person contact attempt (however the vast majority of attempts were in-person).

22. If upon the first contact attempt an enumerator determined an address was occupied and the enumerator was able to obtain a response for the housing unit, then the housing unit was counted, and no follow-up was needed.

23. If upon the first contact attempt, the enumerator was not able to obtain a response, the enumerator was trained to assess whether the location was vacant or unoccupied. Enumerators used clues such as empty buildings with no visible furnishings, or vacant lots, to identify an address as vacant or non-existent.

24. A single determination of a vacant or nonexistent status was not sufficient to remove that address from the NRFU workload; a second confirmation was required. If a knowledgeable person could confirm the enumerator’s assessment, the address was considered vacant or non-existent and no additional contact attempts were needed. A knowledgeable person was someone who knew about the address as it existed on census day or about the persons living at an address on census day. A knowledgeable person

could be someone such as a neighbor, a realtor, a rental agent, or a building manager. This knowledgeable person is known as a proxy respondent.

25. If a knowledgeable person could not be found to confirm the status of vacant or non-existent, use of administrative records could provide confirmation of the enumerator's assessment. The Census Bureau did not rely on a single administrative-records source to determine an address was vacant or non-existent. Rather, multiple sources were necessary to provide the confidence and corroboration before administrative records were considered for use. When used in combination with an enumerator's assessment of vacant or non-existent, corroborated administrative records provided the second confirmation that a nonresponding address was vacant or non-existent.

26. If, upon the first in-person contact attempt, the enumerator believed the address was occupied, but no knowledgeable person was available to complete the enumeration, the Census Bureau used consistent and high-quality administrative records from trusted sources as the response for the household and no further contact was attempted. We consider administrative records to be of high quality if they are corroborated with multiple sources. Examples of high-quality administrative records include Internal Revenue Service Individual Tax Returns, Internal Revenue Service Information Returns, Center for Medicare and Medicaid Statistics Enrollment Database, Social Security Number Identification File, and 2010 Census data.

27. Regardless of whether administrative records were used as a confirmation of vacancy or non-existent status or for the purposes of enumerating an occupied housing unit, the Census Bureau sent, as a final backstop, a final mailing encouraging occupants, if any, to self-respond to the 2020 Census.

28. If a nonresponding housing unit was found to be occupied but no information was gathered on the first attempt, enumerators repeatedly returned. The vast majority of nonresponding addresses in the NRFU workload had the full battery of in-person contact attempts to determine the status of the nonresponding address (vacant,

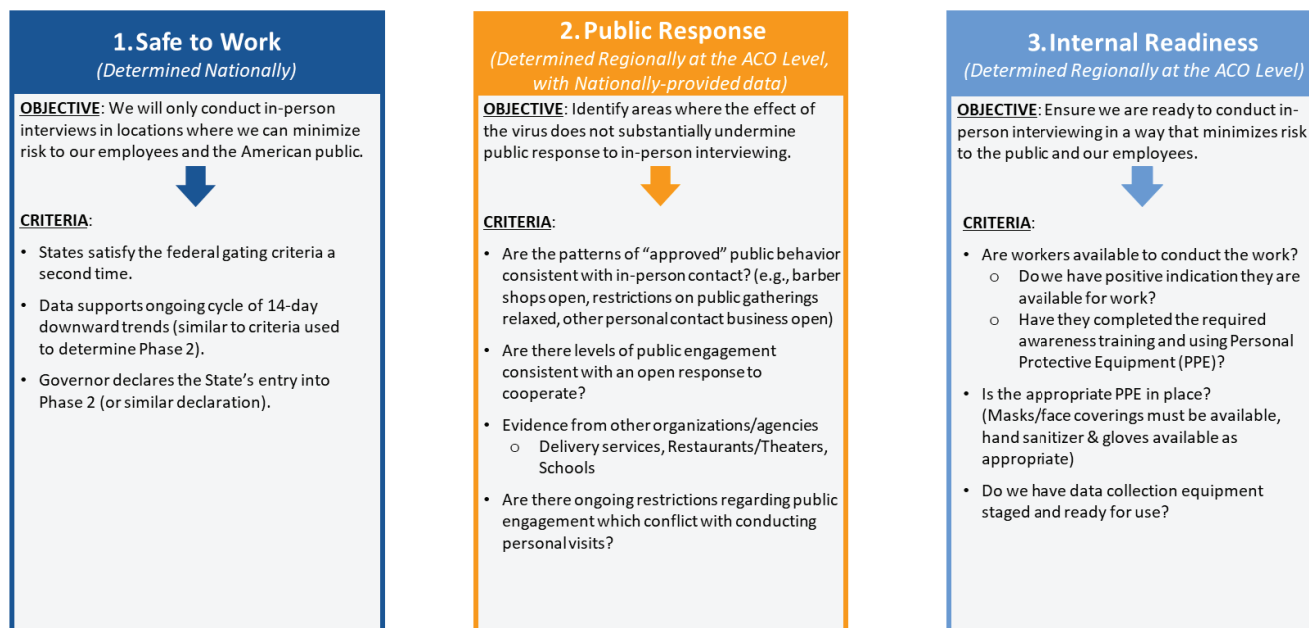
occupied, does not exist) and to collect 2020 Census response data. The full battery of in-person contact attempts also included the ability to collect information from a proxy respondent. Nonresponding units became eligible for a proxy response after three unsuccessful attempts to find residents of a nonresponding address themselves.

29. The Census Bureau arrived at the operational design for NRFU over the course of the decade. Use of administrative records, field management structures, systems, procedures, data collection tools and techniques were proven in tests occurring in 2013, 2014, 2015, 2016, and 2018.

30. While data collection began on schedule, the Census Bureau was forced on March 18, 2020 to announce a suspension of field operations because of the COVID-19 Pandemic. Our original plan was to begin the NRFU operation in most parts of the country in May. But continuing with planned field operations in the spring of 2020 was simply not an option. Many jurisdictions had issued “lockdown” orders. The nation did not know as much about the COVID-19 virus as it does now, and clear public health guidance had not yet been issued. Nor was the Census Bureau able to safely recruit, hire, and train employees for its field operations, and it did not have confidence that households would respond to individuals knocking on their doors seeking responses to the census. Protocols for mask wearing and social distancing were not yet in place and the public health impacts of conducting one of the nation’s largest peacetime mobilizations were unclear.

31. The suspension of field operations and subsequent decisions to adapt field operations were driven by a need to protect the health and safety of the American public; the requirement to implement federal, state, and local regulations on COVID-19; and the desire for a complete and accurate enumeration. We began to re-start operations by resuming our Update Leave operation, resuming pre-NRFU operations in Area Census Offices (ACOs), resuming operations at our paper data capture centers, and resuming fingerprinting and staff onboarding for NRFU workers. The graphic below describes the

criteria we used in our review process for resuming operations during the COVID-19 pandemic.



32. The Census Bureau returned to field operations using a "Soft Launch" approach, meaning that instead of opening all offices at the same time, we instead opened a small number of offices in succession. We opened offices in areas that we believed could be safely started based on COVID risk profiles (developed using CDC, state, and local health guidance), availability of staff, and availability of Personal Protective Equipment (PPE). We needed to acquire PPE, implement social distancing protocols, and work with state and local officials. We opened additional offices throughout the month of July based on detailed daily review of the data about COVID, taking into account state and local stay-at-home orders. We looked for data showing a 14-day downward trend in the area of virus cases, along with sufficient workers to conduct the enumeration, and sufficient available PPE. By August 9 we had begun NRFU in all 248 ACOs. There are 3 ACOs in Alabama. The Census Bureau began NRFU in the Birmingham and Mobile ACOs on August 8, 2020 and commenced operations in the Huntsville ACOs on August 9, 2020.

33. These COVID-19 delays forced the Census Bureau to carry out field operations during hurricane season. Devastating hurricanes in the Gulf Coast area, in particular, limited and slowed the Census Bureau's ability to conduct NRFU operations. Additionally, in large areas of the West Coast, field operations were hampered by conflagrations that caused health alerts due to fire and smoke. And in cities across the country, demonstrations and riots caused further difficulties for in-person enumeration. These challenges forced the Census Bureau to adapt, including by sending travel teams of enumerators to lagging areas.

34. The Census Bureau experienced operational challenges in Alabama in particular. Hurricane Sally hit the state on September 16, 2020 and stopped work in almost all areas of the state for about 3 days. In some places in Alabama, the Census Bureau was unable to work for two weeks. The Mobile ACO was closed for 5 days, from September 15 to 20, 2020. We also had a higher than average rate of COVID-related closures for our ACOs in Alabama. Our practice was to close an ACO for cleaning when workers tested positive for the virus. The Birmingham ACO was closed 4 times (5 days total), Huntsville was closed 3 times (12 days total), and Mobile was closed 1 time (4 days total). When ACOs were closed, we continued field work and operational management via our use of remote technology, but these closures negatively impacted activities such as hiring, training, and payroll.

35. In light of the COVID-19 delays, on April 13, 2020, the Secretary of Commerce and the Director jointly announced a new Census Schedule and stated that they would seek statutory relief from Congress of 120 additional calendar days. This new schedule set a completion date for field data collection and self-response of October 31, 2020. The proposed schedule called for the delivery of apportionment counts to the President by April 30, 2021 (120 days after the statutory deadline) and redistricting data files to the states no later than July 31, 2021. Congress did not pass such a statute.

36. The Secretary and the Director then announced on August 3, 2020 a “Replan Schedule” designed to meet the Census Bureau’s statutory deadline for reporting apportionment data of December 31, 2020. Litigation ensued, and the Census Bureau was enjoined from attempting to meet the Replan Schedule. After a Supreme Court ruling, the Census Bureau ceased data collection operations on October 15, 2020, having resolved 99.9% of all housing units in the process.

Census Step 5: Post-Data Collection Processing

37. Despite the Census Bureau’s best efforts, the delays caused by COVID, hurricanes and wildfires, along with the normal issues commonly encountered during census processing made it impossible for the Census Bureau to finish apportionment processing and deliver accurate and complete apportionment counts before April 30, 2021 (four months after the December 31, 2020 statutory deadline). However, more important in the context of this declaration is that delivery of apportionment data will occur a month later than the statutory due date for delivering *redistricting* data.

38. The order of phases and operations for processing allows no possible way to deliver redistricting data before apportionment data, as one depends and builds upon the other. Further, the complex processing steps that occur between the apportionment delivery include contingency time for rework if it is required, and the delivery of the complete complement of redistricting products cannot be meaningfully shortened or curtailed without unacceptable risk to the accuracy of the data.

39. Below, I generally describe the post processing operations and schedule. While certain steps in different processing operations may sound similar they are in fact quite different because the steps are iterative. For example, we identify unique persons for the purposes of population count in the early phases, while in the later phases we verify and determine demographic characteristics of every unique person.

A. Incorporate address updates from the field data collection operations into MAF/TIGER

Dates: February 6 – September 27, 2020

40. During the data collection operations, the census field staff can update address, update physical location information, and add addresses. These changes are incorporated into our address and geo-spatial MAF/TIGER databases.

41. Once updated, each address must be associated to the correct state, county, tract, block group and block. Since it is critical for many of our data products (including redistricting data) to associate each address to the correct geography, we verify that the address and geo-spatial updates are incorporated correctly.

B. Produce the Final Collection Address Data Products from MAF/TIGER

Dates: September 27 – Oct 14, 2020

42. Once the benchmark has been created, the final collection geographic data files are produced and verified.

C. Produce and review the Decennial Response File 1 (DRF1)

Dates: October 29 – December 26, 2020

43. The verified final collection geography data are integrated with the response data and we verify our work to ensure accuracy. The next set of activities involves the standardization of the collected information.

44. First, we determine the final classification of each address as either a housing units or a group quarters facility, which is necessary because addresses can change from a housing unit to group quarters and vice versa as a result of field observations. Initial status is set at the start of the data collection operations as either a housing unit or group quarters, but during the enumeration operations, we collect information that informs that classification. For a small number of addresses the classification may change, for example a housing unit may have been turned into a small group home.

45. Next we identify persons in housing unit and group quarters returns. As part of NRFU operation, we reinterview a sample of cases to ensure quality. We incorporate the results of the reinterview for further action in subsequent steps.

46. As part of the Internet self-response option and telephone operation, respondents can provide their data without their Census Identification Number (ID). These cases are assigned an ID which associates them to the final collection geography.

47. We collect data in many ways in the Group Quarters operation, for example on-line, over the phone, on a paper questionnaire, electronic administrative files, and in person using an electronic questionnaire. As a result, we need to standardize the responses across the modes of collection. This step ensures all the data are formatted in the same way for subsequent processing.

48. Finally, for the operations that collect data on a paper questionnaire, some housing units have more people than can fit on one paper questionnaire. The census field staff will use multiple paper questionnaires to enumerate the house. We have to link these continuation forms to form one household.

D. Produce and review the Decennial Response File 2 (DRF2)

Dates: December 26, 2020 – February 26, 2021

49. Once the previous step has been verified, we incorporate the results from the Self-Response Quality Assurance operation that occurred during the data collection phase. This quality assurance operation ensures that data determined to be falsified or incorrect are not added to the Census.

50. As part of the group-quarters operations—a special enumeration procedure used for group housing, such as prisons and colleges—we enumerate domestic violence shelters. Their locations and data are highly sensitive and are handled with special procedures both in the field and in processing. Their data are incorporated at this point in the process due to their sensitivity.

51. Also at this stage in the process, we select a form that will be used as the enumeration of record for those small number of addresses where we receive multiple returns—for example, where one person in a house completes the form online, and another completes the paper questionnaire. This is another important step to avoid duplication, and was particularly important for the 2020 Census, given the multiple modes of response that were offered and the ability to respond without an assigned identification number.

52. As with the prior steps, we continue in the DRF2 file to identify, review, and resolve data anomalies to ensure that the data are accurate.

E. Produce and review the Census Unedited File (CUF)

Dates: February 27, 2021 – March 10, 2021

53. It is in the CUF that we began to incorporate administrative records data as the response data for housing units where we do not have an enumeration but where we have high-quality administrative records data. Incorporating this information thus helps the Census Bureau achieve a more complete census. Administrative record data can include information previously collected by other federal or state agencies, including the Internal Revenue Service and the Social Security Administration.

54. Next, we finalize the status for every housing unit as occupied, vacant or non-existent. Non-existent units are removed from future processing. For every occupied housing unit, we determine the population count.

55. For every housing unit and group quarter, the location is processed by state, county, tract, block group, and block. Then we verify the status (occupied, vacant or non-existent) for every housing unit and group quarter. And in all occupied addresses, we verify the number of persons. For a small number of individuals who live in transitory locations such as campgrounds and marinas, this information will be added later in the process.

56. For unresolved housing units—*i.e.*, those we believe to be occupied but do not have respondent data after all collection operations end—we use a statistical method called count imputation to assign a population count. This was done for less than one half of one percent of housing units in the 2010 Census.

57. The result of these processes is a file that contains records for every housing unit and group quarters along with person records for the people associated with the addresses, although some of the demographic and tenure information may still be missing by this step.

58. As of the date I execute this declaration CUF processing is complete.

59. All of the processes involved in CUF creation are critical predecessors to the subsequent processing operations, including those needed for redistricting.

F. Produce, review and release the Apportionment Counts

Scheduled Dates: March 12 - April 30, 2021

60. On completion of the CUF, we verify and incorporate into the state population counts data that includes the Federally Affiliated Overseas population (such as military personnel serving overseas), the results of the Enumeration of Transitory Locations for each State, and any addresses added late in census data collection.

61. Next, we determine the apportionment counts. Since all housing units and group quarters have a population count linked to a State, we can feed their tabulation into the state-level population counts.

62. To ensure accuracy in the apportionment numbers, the state counts including the overseas population and apportionment numbers are verified by independent teams using different methods. The results of the independent verifications are compared and reconciled, if necessary.

63. We produce the apportionment results using the method of equal proportions. We deliver the apportionment package to the Secretary of Commerce who then delivers them to the President. The President then reports the numbers to Congress.

64. Once again, the Census Bureau does not believe it can deliver apportionment counts sooner than its current working schedule.

G. Produce and Review Census Edited File (CEF)

Scheduled Dates: April 20 – June 24, 2021

65. The next stage is the creation of the Census Edited File (CEF). Whereas the CUF was the basis for the apportionment counts, the CEF provides for the much-more detailed and voluminous data required for redistricting. This process adds any remaining missing values, rectifies conflicting information, and overall ensures a complete set of records for subsequent production of redistricting data.

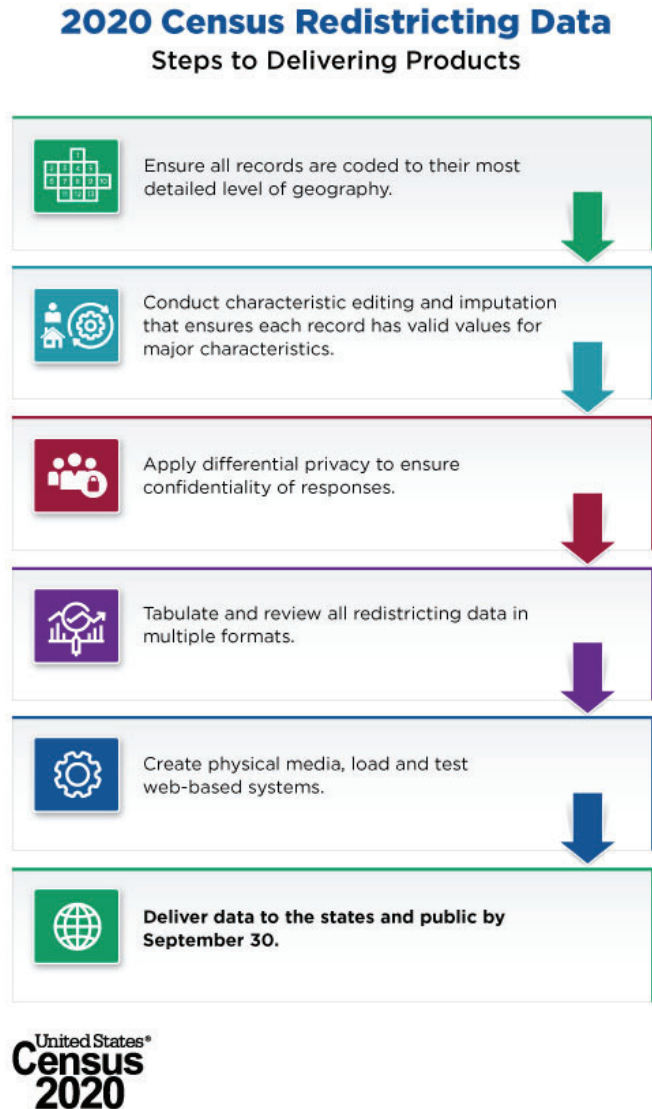
66. It is at this stage that the detailed information about individuals living in the households is compiled (as opposed to the simple population count). This includes complex processing for the race and ethnicity and age information that states must have in order to conduct redistricting.

67. While processing for apportionment only requires accurate population counts, the detailed respondent information needed for the CEF can be conflicting or contradictory and requires application of complex editing rules. Additionally, missing data are accounted for using a statistical process called characteristic imputation.

68. This is a highly complex operation that involves iterative run and review cycles to ensure processing is occurring as designed. And as anomalies are uncovered and corrected, the data must be re-processed.

69. Due to the complexity of the operation, and the iterative cycles of review, it is not possible to accelerate the production of the CEF without unacceptable risks to data quality.

70. The general steps for production of the redistricting data are depicted in the following graphic.



H. Produce and Review Microdata File (MDF)

Scheduled Dates: June 25 – July 18

71. The next stage of the process is to create the privacy-protected Microdata Detail File. At this stage the Census Bureau applies formal privacy protections to prevent

revealing respondent information. This stage is expected to take three weeks; creation of the MDF is not the reason that the Census Bureau will be unable to meet the statutory deadline for delivering redistricting data. In fact the disclosure avoidance procedures completed in the 2010 census processing took 27 days – or nearly four weeks.

72. Application of these protections is not optional. The Census Bureau is required by law to protect the confidentiality of the information provided by respondents. Consistent with that requirement, the Census Bureau designed the 2020 Census Disclosure Avoidance System (DAS) to ensure equal privacy protections for every enumerated person in the country, irrespective of where they live, and that the accuracy of any statistic that we publish will improve as the number of people being measured increases. Our disclosure avoidance approach – differential privacy – accomplishes these objectives by taking detailed privacy-protected measurements of the population at all levels of geography, from the nation down to the individual Census block, and processing these data in descending order from the nation down to the individual block. This method is discussed in greater detail in the declaration of John Abowd.

73. Because of how the method works, the disclosure avoidance algorithm must be applied to the full census data set (i.e., the entire nation) in order to function correctly. Processing these measurements in a top-down fashion, starting at the national level, allows the algorithm to improve the accuracy of the statistics at all geographic levels without impacting the privacy guarantee by leveraging the accuracy of statistics at higher geographic levels. This stage also requires careful review – and, if necessary, re-processing of the data – to ensure the system is functioning as designed.

74. As explained in the declaration of John Abowd, if the Census Bureau were ordered to adopt a different disclosure avoidance methodology at this point in time, it would add significant additional time (at least several months) to the schedule for delivering redistricting data. We would be required to develop an entirely new DAS system, including developing and testing new software. Switching to a new DAS system at the

eleventh hour would also pose significant risks to data accuracy given that the alternative systems (swapping and suppression) are blunt instruments that, unlike differential privacy, cannot be effectively tuned to optimize for data accuracy.

I. Produce and Review Tab file

Scheduled Dates: July 19 - August 16

75. Next, the Census Bureau conducts the tabulation and review of the tabulated census data. Tabular data are easily understandable and usable data tables that the public expects from the Census Bureau. Prior to tabulation, the census data are still in a largely unusable form that would require significant expertise and manipulation on the part of data users to select and understand useful information. Tabulation (literally, formatting and summarizing data into ‘tables’) makes census data easily accessible by state officials or the public. Published tables must be created from the processed data; must be accurate and complete; and must integrate geography, population, and characteristics in myriad combinations.

76. The voluminous and detailed nature of tabulation requires rigorous review and validation that cannot be skipped or shortened without significant risk to the accuracy of these products. This is because tabulation products are created by further processing the base data. The expert review ensures each tabulation product aligns correctly with the base data.

77. As noted in the declaration of James Whitehorne, the Census Bureau announced on March 15, 2021 that it would make a legacy version of redistricting data available to the states in mid-to late- August. While the legacy version requires more data processing expertise on the part of the states, it will be accurate, privacy-protected, and fully usable for redistricting purposes.

78. Finally, during tabulation we also add new fields in the data that make future extraction, summation, and deeper understanding easier for data users to achieve. For example, we may need to add a “voting age” field and update our system so that all

records for people age 18 or over have the that value. This enables “filtering” for the characteristic of “voting age.”

79. If the Census Bureau were to prioritize one State’s legacy version of redistricting data with all planned user tools (to the detriment of the other 49 states), it may be able to deliver a few days earlier than other States, at most.

J. Produce, Load, and Disseminate Redistricting Data

Scheduled Dates: August 17 – September 30

80. The final stage is the review, preparation, loading, and delivery of the official redistricting data. This stage involves the creation of dissemination materials to send the states, the loading of web-based systems, the testing of those materials and systems to ensure they are functioning correctly, and the actual delivery of the data to the states and the public.

81. After we test our materials and systems to ensure that they are functioning correctly, we then create redistricting materials to send to the States. In order to provide convenient access to data users, we also load these data products to our web-based systems. And we deliver the data to the states and the public.

82. While this is happening, we are loading the data.census.gov Data Explorer tool with the entire nation’s data and metadata that allows the system to properly pull and display data. This too must be carefully verified to ensure that data for every geographic level and every table is being properly pulled by the application and displayed. Once all of these materials are ready for release, the physical materials are mailed to the official recipients via overnight mail and the web-tool is made available to the official recipients and the public.

83. If the Census Bureau were to prioritize the DVD/Flash Drive and the data.census.gov webpage for one State’s redistricting data (to the detriment of the other 49 states), it would not be able to deliver the data more than a few weeks earlier than a single national release.

Impossibility of Producing Tabulated Data Prior to September 2021

84. As explained above, the Census Bureau requires approximately five months from the release of the apportionment data on or around April 30 to produce and review the Census Edited File, the Microdata File, the Tab file, and then ultimately produce the redistricting data by September 30. Although the 2020 Census Operational Plan provided for only three months from the planned release of apportionment data of December 31, 2020 to the planned release of redistricting data on March 31, 2021, the Census Bureau now requires an additional two months because of operational changes that the Census Bureau made to expedite the release of the constitutionally required apportionment counts.

85. Specifically, in order to ensure the release of the apportionment counts as quickly as possible, the Census Bureau “decoupled” certain processes that the Census Bureau would have normally completed at the same time. For example, processing of the Census Unedited File (CUF) had pieces set aside to concentrate on processing operations that focused only on population counts for apportionment. This decoupling required designing and creating a second, later CUF format that enabled accurate processing for not only population counts, but also for the demographic characteristics required for redistricting. Under the 2020 Census Operational Plan, the Bureau would have completed this work in parallel with the work on the apportionment counts. This second Census Unedited File would not have been necessary if the Bureau had not, of necessity, decoupled these steps.

86. The delay that has resulted from this “decoupling” of certain processes to prioritize the release of apportionment data accounts for much of the added time in the schedule. The remainder of the additional time accounts for the likelihood that the Bureau will encounter additional “anomalies” that it will need to review and resolve. While it is conceivable that the Bureau may be able to release redistricting data a few weeks earlier if there are zero anomalies, in the Bureau’s experience, this is highly unlikely.

However, if the Bureau is able to produce redistricting data earlier, it will of course do so.

Commitment to Data Quality and Transparency

87. The Census Bureau is committed to unprecedented transparency about the quality of the 2020 Census. The Deputy Director set up a Data Quality Executive Guidance Group (EGG) in April 2020 to evaluate the quality of the 2020 Census. The EGG draws upon expertise of career employees within the Census Bureau in the fields of census operations, statistical methodology, acquisition and use of administrative records, and in the social, economic, and housing subject areas. The group has been meeting continuously since that time and fully supports of the processing steps and the timeline described above to ensure that 2020 Census data are accurate and complete.

88. The Census Bureau is now collecting data for its formal coverage measurement study, the Post-Enumeration Survey. That effort will provide estimates late in 2021 of coverage errors (both undercounts and overcounts) for States and for various demographic groups, similar to what we released after the 2000 and [2010 Censuses](#).

89. In December 2020, the Census Bureau released the 2020 Demographic Analysis (DA) estimates. DA estimates consist of national-level estimates of the population by age, sex, race, and Hispanic origin as of April 1, 2020. These estimates are developed from current and historical vital statistics: birth and death records, estimates of international migration, and Medicare records. The Demographic Analysis estimates are independent from the 2020 Census and are used to calculate net coverage error, one of the two main ways the Census Bureau evaluates the coverage of the census.

90. The 2020 Census has been the most challenging census in modern history. The Census Bureau has faced an unprecedented pandemic, natural disasters, and civil unrest in addition to the already-complicated task of collecting and processing data for nearly 150 million households across 3.8 million square miles. Despite these challenges, the Census Bureau resolved 99.9% of all housing units in the nation and it has planned a


schedule that is designed to achieve the complete and accurate data that will guide the country for the next ten years. At all points, delivery of complete and accurate data has been our overriding priority.

91. I respectfully submit that the intentions of the dedicated professional staff at the Census Bureau will always be to deliver accurate data about the U.S. population and economy, and we continue to do so as we work through the challenges we as a nation have faced over the last year.

I have read the foregoing and it is all true and correct.

DATED and SIGNED:

**MICHAEL
THIEME**

 Digitally signed by MICHAEL
THIEME
Date: 2021.04.11 21:00:34
-04'00'

Michael Thieme

Assistant Director for Decennial Census Programs, Systems, and Contracts

United States Bureau of the Census

**IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA**

STATE OF ALABAMA, *et al.*,

Plaintiffs,

v.

UNITED STATES DEPARTMENT OF
COMMERCE, *et al.*,

Defendants.

Case No. 3:21-cv-211-RAH-ECM-KCN

DECLARATION OF JAMES WHITEHORNE

I, James Whitehorne, make the following Declaration pursuant to 28 U.S.C. § 1746, and state that under penalty of perjury the following is true and correct to the best of my knowledge and belief:

1. I am the Chief of the Census Redistricting and Voting Rights Data Office at the U.S. Census Bureau. I have occupied this position since July 2015. Prior to that, I served as the Assistant Chief in the same office from April 2010 until becoming Chief. As Chief of the Census Redistricting and Voting Rights Data Office I am responsible for management of the Census Bureau's redistricting data program and for implementation of 13 U.S.C. § 141(c). I am knowledgeable about the Census Bureau's redistricting data program.

2. I am making this Declaration in support of Defendants' Opposition to Alabama's preliminary-injunction motion. All statements in this Declaration are based on my personal knowledge or knowledge obtained in the course of my official duties. In this declaration I:

- Provide background on the Census Bureau's redistricting data program;
- Explain the process by which the Census Bureau established September 30, 2021 as the working schedule date by which we would complete delivery of redistricting data, and our reasons for establishing this schedule;
- Explain why it is impossible for the Census Bureau to comply with the statutory deadline set in § 141(c); and
- Explain the likely effect of any order compelling production of redistricting data for Alabama prior to the completion of processing.

Background on the Redistricting Data Program

3. Section 141(c) of the Census Act requires the Secretary of Commerce ("the Secretary") to establish a program allowing States to identify the geographic areas for which specific tabulations of population are desired. Section 141(c) also directs the Secretary to deliver basic tabulations of population, and geographically specific tabulations

for those States participating in the program, to the Governor and officers or public bodies having responsibility for legislative apportionment or districting within one year from the decennial census data (which is April 1).

4. The Census Bureau established the program after passage of Public Law 94-171 in 1975, codified at 13 U.S.C. § 141(c). The States generally use redistricting data produced under § 141(c) to redistrict for state and congressional elections, although they are not required to do so by any federal law. The U.S. Department of Justice also uses redistricting data to enforce of the Voting Rights Act.

5. Section 141 requires the Census Bureau to conduct the program in a non-partisan manner. We accomplish this by asking each State to assign a non-partisan liaison or liaisons at the start of the program each decade. The redistricting data program asks that the majority and minority leadership in all chambers of the state legislatures sign off on the individuals they feel can represent the State in a non-partisan manner. The redistricting data program then works with those people (or their successors) for the lifecycle of the program.

6. The redistricting data program (RDP) works to ensure the states are informed about the decennial census and the RDP. We started the 2020 RDP by offering in-person briefings to each state, eventually providing information about the 2020 Census and the 2020 Redistricting Data Program to the 26 states that accepted. We have continued to conduct state briefings when requested. We keep as many states as possible informed through our regular interactions with umbrella organizations such as the National Conference of State Legislatures. We also interact directly with our program liaisons in each state.

7. When the Census Bureau first requested a four-month statutory extension from Congress in April 2020, we called our liaisons in New Jersey and Virginia because those two states require redistricting data prior to our statutory deadline. We also emailed all of our liaisons over the month of May 2020 to try to understand the impact of

that extension request. Some States were able to act on that information, such as New Jersey where voters approved a constitutional amendment that allowed the State to use previous district maps until the new maps are in effect for the 2023 elections. And in California, the state legislature sought and obtained at least a four-month delay of its redistricting deadlines from the California Supreme Court.

8. More recently, we notified all of our liaisons on January 28, 2021 that the redistricting data would be delivered later than July 31, 2021, and we have been providing a direct line of support for the states and our liaisons when they come to us with census-related questions. For example, we worked with the Secretary of State's office in Idaho to help them identify data that will allow them to perform a series of initial draft plans, thereby reducing their workload by 50% when the official data is provided. *See [Webinar, Comments by Jason Hancock, Deputy Secretary of State, Idaho \(March 5, 2021\)](#).*

9. The 2020 Census redistricting data program is being conducted in five phases. The first two phases are the Block Boundary Suggestion Project and the Voting District Project. These two phases were conducted in advance of the decennial census in the years 2015 through 2020 to provide States the opportunity to identify the geographic areas for which specific tabulations of population are desired. The third phase of the program is delivering redistricting data to the states. In the fourth phase, the Census Bureau collects the newly redistricted congressional and state legislative districts created by the States after the Census Bureau delivered the redistricting data. This phase is also used to collect changes every subsequent two years if States redistrict again during the decade. In phase five, the Census Bureau evaluates the previous decade's program, incorporating feedback from the States, and develops an outline for the next decennial's redistricting data program.

10. Currently, the Census Bureau is in phase three, the data delivery phase. We provided the States with geographic support products in January and February of this year. State and local governments use these products in their redistricting efforts. The

products contain newly created 2020 Census blocks and updated block groups, census tracts, voting districts, and current boundaries for legal governments and school districts referenced to January 1, 2020. Using the information that each State provided, we have now delivered the geographic information that will help them plug in the actual 2020 Census data and do their work of redrawing district boundaries.

11. States can use these supplied geographic files now to prepare for redistricting and elections administration. In some states, there is a requirement that incarcerated individuals be reallocated from the correctional institution where they are counted by Census to a pre-incarceration address. With the release of the geographic support products, states can identify the location from and to where these individuals need to be moved. This adjustment can be prepared in advance at the granular census block geography and then those adjustments can be applied to the data upon its release.

12. States, including Alabama, can do the same to prepare voter rolls. The Census Bureau provided both geographic information system (GIS) files and .pdf maps for printing that clearly show every block in every county in every state. Both the GIS files and the .pdf maps include information on roads and other geography. That information can be used now to identify in which census block each voter resides. If the state does not split census blocks when preparing their redistricting plans, as most states do not, then it would be a simple matter to associate those census blocks with the new plans after they are drawn to recreate voter rolls for the new electoral geography. If the state were to split census blocks, then the same association can be made for the majority of blocks kept whole and with minimized rework for those blocks that are split.

13. We are now preparing to deliver the official data, once it has been processed and cleared for publication. This data will be delivered to the States in two methods. The first method is on DVDs and Flash Drives. These physical devices will have an integrated software browsing tool that will allow intuitive browsing of the data. They also contain a custom extraction menu that allows for the extraction of large datasets from the device.

Those extractions can then be imported easily into a Geographic Information System or database. The second method is using our data.census.gov webpage. This webpage is a data browsing tool where data users can access many different census datasets, including the redistricting data. It has custom filters that allow the user to filter on those geographic and characteristic data for which they are interested. For example, a State could filter the data and easily identify the number of voting-age residents by race or ethnicity in each and every block within a census tract, county, or even for the entire state. Data users can view, map, and download these datasets once they have set the filters with their choices.

Impossibility of Complying with the Statutory Deadline

14. The Census Bureau has not yet finalized or produced the redistricting data as of the date I sign this declaration, which is past the statutory deadline of March 31, 2021. Based on my knowledge of decennial census data processing, it is not possible under any scenario for the Census Bureau to produce these data at this time or any time in the immediate future, and the Census Bureau would be unable to comply with any such order from the Court. Simply put, it would be a physical impossibility.

15. As explained in depth by the Assistant Director for Decennial Census Systems and Contracts, Michael Thieme, the Census Bureau must complete a series of interim steps prior to delivering the redistricting data:

- On completion of the initial CUF, we incorporate the Enumeration of Transitory Locations data, and any addresses added late in census data collection into the detailed final version of the CUF, scheduled to be completed by April 19, 2021.
- The Census Edited File (CEF) in the working plan is scheduled to be completed by June 23, 2021. To produce the CEF, the final CUF needs to go through the editing and imputation process which ensures all records have valid values.
- The Microdata Detail File (MDF) in the working plan is scheduled to be completed by July 17, 2021. Census data is protected by Title 13 and cannot be disclosed until Census completes disclosure avoidance processing. The output of those privacy protections is

the MDF. Disclosure avoidance involves a privacy-loss budget based on complex algorithms that requires the entire national set of CEFs as its input. In other words, it is impossible for this step to be completed until CEF data from all states are done processing.

- The tabulation system then uses the MDF to append more detailed geographic information and then generate the tabulated versions of the data in all of their needed formats. These formats include the text-based summary files and the Application Programming Interface format that drive the data.census.gov website. These formatted tabulations are then reviewed by subject matter experts to ensure the tabulations were performed correctly. In the working plan, this review completes on August 13, 2021.
- The final activity is the production, loading, and testing of the actual dissemination materials and system. The materials are DVDs/Flash Drives with custom browsing software to make accessing the data user friendly. This requires another format conversion of the tabulated data. The system is the data.census.gov platform that provides access to the data for the states and the public. These materials need to be created, system loaded, and all reviewed and tested prior to being provided to the states.

16. Each of these interim steps, in order, is required to move to the next. And the processing for each of these interim steps are interrelated, so changing something in one would impact each subsequent step and threaten the success of the overall process. In the working plan, the Census Bureau is scheduled to complete production of redistricting data no later than September 30, 2021.

17. The current working schedule for producing the redistricting data is not set in stone, however. The schedule builds in time to account for multiple reviews on the assumption that anomalies will be discovered and must be corrected prior to publication. Anomalies found in processing are not errors in the census, but they can turn into errors if we don't review and resolve them. In a perfect world where every single step of the processing occurs with no additional anomalies or impediments, it may be possible to deliver redistricting data weeks before September 30. But it has been our experience with

the 2010 Census processing, with the first half of the 2020 Census processing, and with the issues we already know we will encounter in the second half of the 2020 Census processing, that there is little to no chance of encountering this perfect world. This is a dynamic process and the Census Bureau is doing everything in its power to produce high-quality redistricting data as quickly as possible.

The September Delivery Date

18. The Census Bureau's original plan was to release the redistricting data in a staggered fashion, releasing a group of states each week between February 18, 2021 and March 31, 2021. But by early 2021 it was clear that the Census Bureau could not meet its statutory deadline for delivering redistricting data to the states. Around the same time, my office received questions from our state liaisons about when we could deliver the data, and we wanted to provide this information to the States so that they could plan for the delay. The Census Bureau therefore sought to establish an achievable schedule for redistricting data that built in sufficient time for review and revision, and produced redistricting data that States could use with confidence.

19. To create the current working schedule, I coordinated with many different components of the Census Bureau about the time each office needed to perform its part of the process. These areas include: Decennial Response Processing System, Decennial Statistical Studies Division, Economic Reimbursable Surveys Division, Production Environment for Administrative Records Staging Integration and Storage, Demographic Directorate, Disclosure Avoidance System, Tabulation System, Center for Enterprise Dissemination Services and Consumer Innovation, Application Development and Services Division, Decennial Census Management Division, and the Census Redistricting and Voting Rights Data Office.

20. Recognizing that processing schedules are a snapshot based on past and current experiences, the working schedule that we created—estimating completion by September 30, 2021—takes into account the Census Bureau's processing experiences thus

far (such as the discovery of anomalies) and allows time for subject matter review and reruns of files, if necessary. To the fullest extent possible, it adjusts the sequencing and durations of some operations to achieve efficiencies. For example, the working schedule has allowed us to prepare ancillary files needed for creation of the CEF, while waiting for receipt of the CUF. Originally, these were to be processed after receipt of the CUF. This change allowed us to “save” several weeks in expected processing time.

21. But as discussed above and by Mr. Thieme, this highly complex operation involves iterative and interrelated review cycles to ensure processing is occurring as designed. This is crucially important because the finished CEF becomes the source of all decennial data for the next ten years. After completion and validation of the CEF, it is also essential that the Census Bureau process the country as a nation through the disclosure avoidance process to protect the confidentiality of all census respondents. These two necessities (completion of CEF and privacy protections) in particular push us deep into the planned schedule, well beyond the statutory deadline.

22. In addition, critical decisions are made using the data produced by the Census Bureau for the decade following their publication. The processing work that leads to the redistricting data products ensures the eventual quality of not only the redistricting data but all of the major decennial data releases that are used throughout the decade for public policy, funding formulas, business decisions, and many other uses.

23. To help states and the public plan, on February 12, 2021 the Census Bureau announced the new working schedule and published a blog (available [here](#)) that I had written describing our process and rationale.

24. Originally, we planned a staggered delivery so that we could order states based on their redistricting deadlines, prioritizing states that needed the data sooner. In our efforts to keep the redistricting schedule as short as possible while maintaining the quality of the data, however, we determined that a single national delivery would pro-

vide an overall shorter timeframe than a staggered release. That's because a single national release will enable us to compress several production and review activities. For example, by moving to a single national release, DVD/Flash Drive creation and review went from 37 days to 28 days; review of the tabulated data went from 36 days to 20 days; and the load-and-review process for the data.census.gov data website went from 42 days to 23 days.

25. The single national release will also allow the Census Bureau to ensure the delivery of redistricting data with finality, possibly saving time over a staggered release. When performing data reviews, there may be an error in one State that is not apparent except when viewed in another State or multiple States. These findings act as triggers to perform additional reviews to identify whether it is a systematic error that may require reprocessing of all States' data or if it only affects the State in which it was found. So a single national release allows the Census Bureau to complete the review of all the dissemination materials prior to release, thus reducing the likelihood of finding an error after the data for one State was released that would require us to retract that data, conduct additional processing, and reproduce that State's corrected data much later. Because the Census Bureau will be unable to meet the statutory deadline to provide the redistricting data to the States, and with the urgency of supplying all States their data as soon as possible, it is thus more efficient for all States' data to be reviewed prior to dissemination.

26. The Census Bureau is aware that there are now a number of States, including Alabama, that will have to address statutory or state constitutional issues resulting from our delayed delivery of the redistricting data. Based on the National Conference of State Legislature's webpage titled [*2020 Census Delays and the Impact on Redistricting*](#), at least 27 states have a requirement for redistricting to be completed in 2021 (either explicitly or implicitly). And each state has its own constitutional and statutory requirements, some of which include public meetings, data modification, and other requirements. So,

with the delay in the delivery of the redistricting data, there are now too many states (at least 27) to prioritize, in a fair, logical, and data-driven manner.

Release of Legacy Format Redistricting Data File in August

27. In recognition of the difficulties the above timeline creates for states with redistricting and election deadlines prior to September 30, the Census Bureau continuously reviews its timeline to identify any opportunities to shorten the processing schedule. Our review confirms that all steps of data processing and formatting will be complete by September 30. However, we recently announced on March 15, 2021, that we expect to be able to provide states with a “legacy” format summary redistricting data file in mid-to-late August. The legacy format summary files will not contain the individual data tables that will be included in the September release, but states can use an outside vendor to process the data if they do not have the capacity to extract individual data tables from the legacy format data on their own. While we had intended to provide the legacy format summary files with the final 2020 Census redistricting data, we determined that states should be given the opportunity to use the legacy format files as soon as they become available in August. The legacy format files will have identical data to the files that we expect to deliver in September. They will have been fully reviewed and subject to the same exacting quality assurance processes. The only drawback to using the legacy format summary files is that they will require additional handling and software to make the data accessible. We expect that many states will elect to use the August delivery because they have used similar products in the past and the vendors who support the states in redistricting are also knowledgeable and able to work with these files.

28. The product and integrated tools we plan to deliver in September—DVD/flash drives and our data.census.gov Data Explorer platform—will have enhancements that make accessing and working with the data more user-friendly. The DVDs and flash drives will have an integrated software browsing tool that will allow intuitive browsing of the data. That software also contains a custom extraction menu that

allows for the extraction of large datasets from the device. The data.census.gov Data Explorer platform is a data browsing tool where users can access many different census datasets, including the redistricting data. It has custom filters that allow the user to filter on those geographic and characteristic data in which they are interested. For example, a state could filter the data and easily identify the number of voting-age residents by race or ethnicity in each block within a census tract, county, or even for the entire state. Data users can view, map, and download these datasets once they have set the filters with their choices.

Prioritization of States

29. Alabama is not the only state that has sued the Census Bureau seeking earlier delivery for its redistricting data and asking that it be prioritized. If the Census Bureau were to prioritize the DVD/Flash Drive and the data.census.gov webpage for one State's redistricting data (to the detriment of the other 49 states), it would not be able to deliver the data more than a few weeks earlier than a single national release. If the Census Bureau were to prioritize the older, more-complicated file format for one State's redistricting data (to the detriment of the other 49 states), it may be able to deliver that data a few days earlier than other States, at most. As explained above, the Census Bureau cannot produce data for any State until after the disclosure avoidance (privacy protections) have been applied, which requires processing data for all States at once. As a result, even if the Census Bureau prioritized the redistricting data for one State, it could only focus on that State after privacy protections (i.e., disclosure avoidance) are applied, and we would still need to create and review the data tabulations thereafter. However, if one State were prioritized through those reviews, the resulting data may have uncaught errors from being rushed through review without the benefit of review of all States at once, perhaps ultimately sacrificing both accuracy and time (as discussed above).

30. Prioritizing one state would also divert the use of resources and systems that are needed for the national release and delay the release of data for the other 49 states.

That's because prioritizing one State would mean focusing the Census Bureau's resources on that review and processing, delaying review for all other States. So even if the prioritized one State's data could advance the DVD/Flash Drive and data.census.gov webpage release by several weeks, the same data for the other 49 States would be delayed.

31. The Census Bureau's working schedule does not contemplate advantaging one State over the other 49. So, the full extent of any additional delays resulting from the prioritization of one State over the other 49 would have to be determined by recreating the working schedule with the one prioritized State ahead of all others. Based on my current knowledge, attempting to prioritize one State may cause an additional delay by as much as several weeks, causing even further disruptions for the remaining 49 States' redistricting processes.

32. The current situation is understandably frustrating to Alabama, and to the majority of States. As the officer within the Census Bureau charged for advocating on behalf of the States in regard to the redistricting data program, I understand and share their frustration. But dedicated Census Bureau professionals are working as diligently and efficiently as possible to ensure that the data we provide for redistricting are produced as quickly and as accurately as we can accomplish.

33. I have read the foregoing and it is all true and correct.

DATED and SIGNED:

JAMES WHITEHORNE

Digitally signed by JAMES
WHITEHORNE
Date: 2021.04.11 18:49:59 -04'00'

James Whitehorne

Chief, Census Redistricting and Voting Rights Data Office
United States Bureau of the Census